

*Basics in Proteomics*

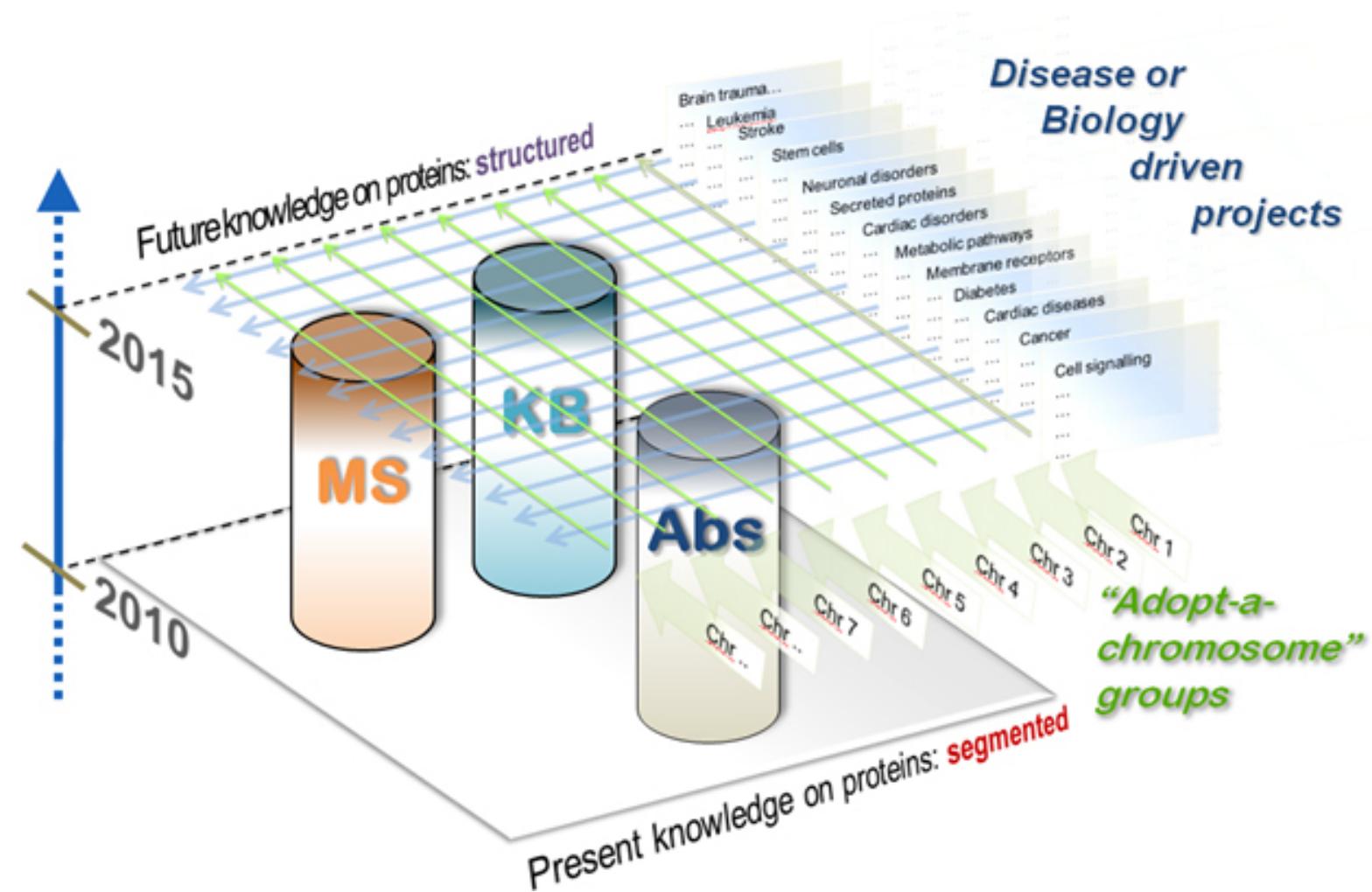
&

*Biomarker Discovery*

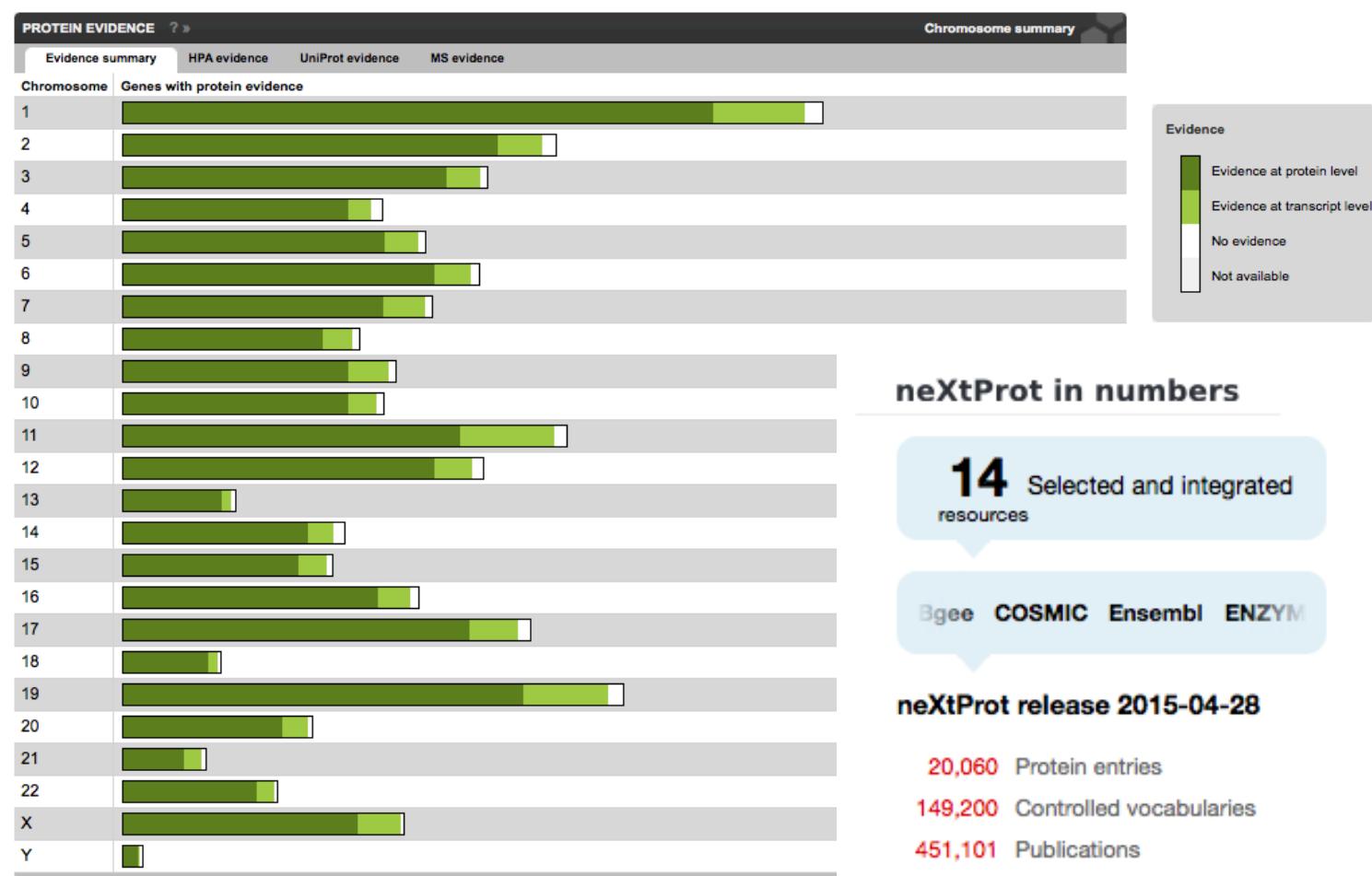
PART 1

# **BASICS IN PROTEOMICS**

# THE HUMAN PROTEOME PROJECT

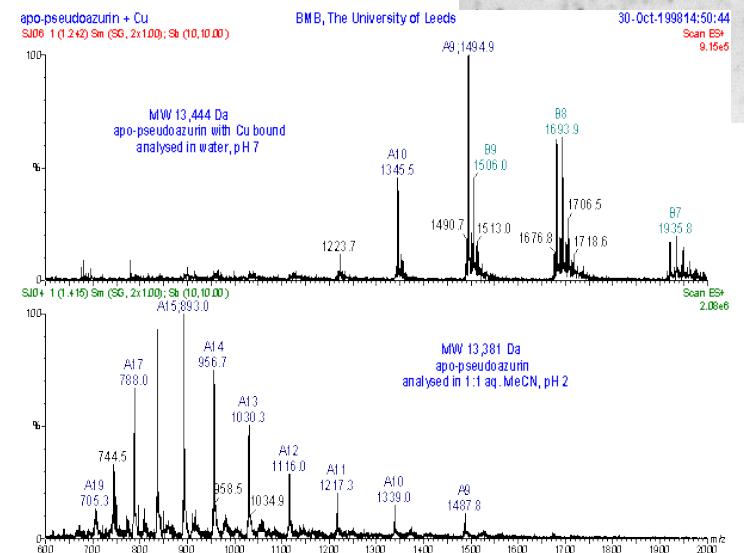
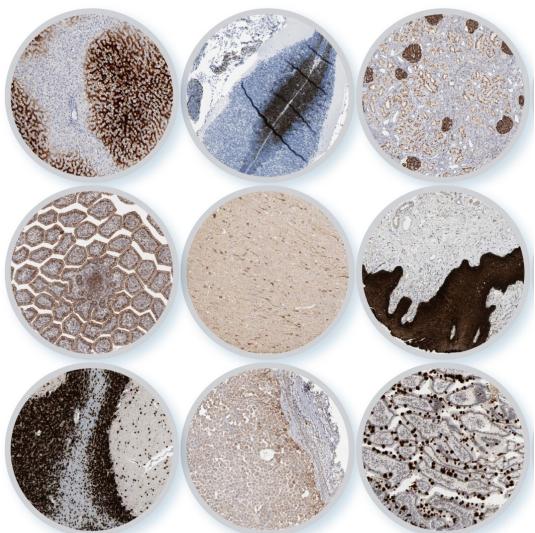


# MAPPING THE HUMAN PROTEOME

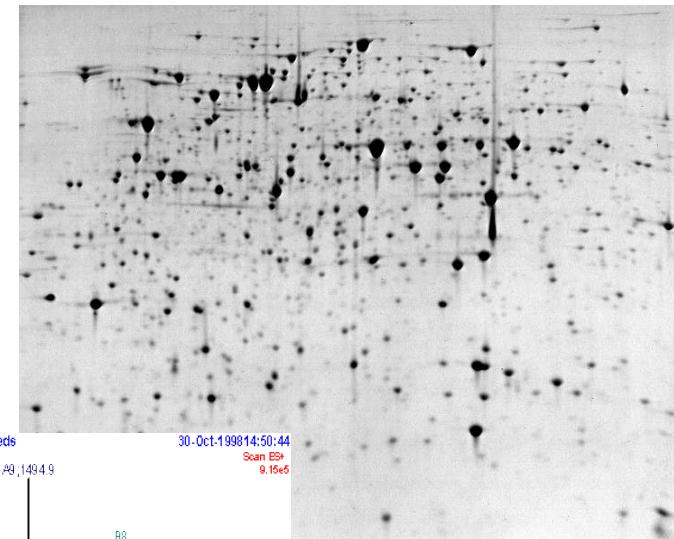


# PROTEOMIC TECHNOLOGIES

- Separation-based
  - Mass-based
  - Affinity-based



Taken from - Dr Alison E. Ashcroft, The University of Leeds.



Taken from - [www.biopoint.co.u](http://www.biopoint.co.u)

# THE CHALLENGES OF PROTEOMICS

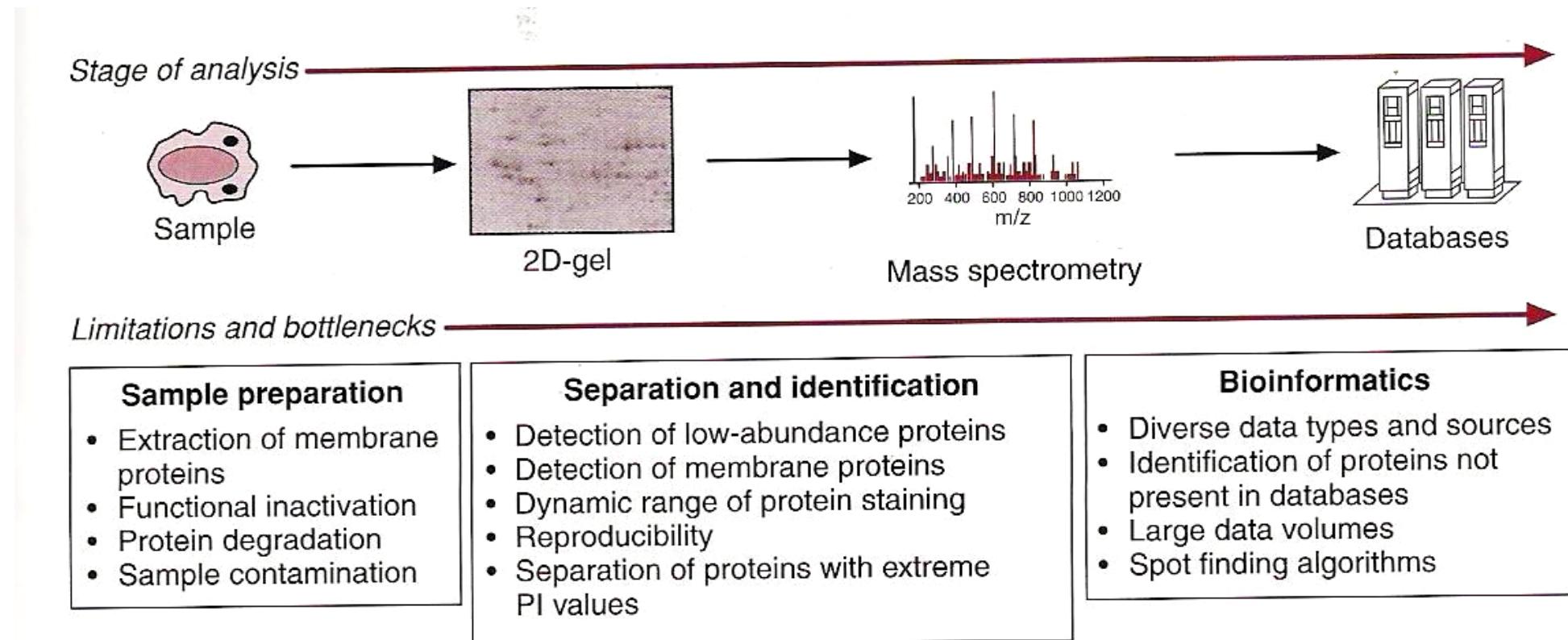
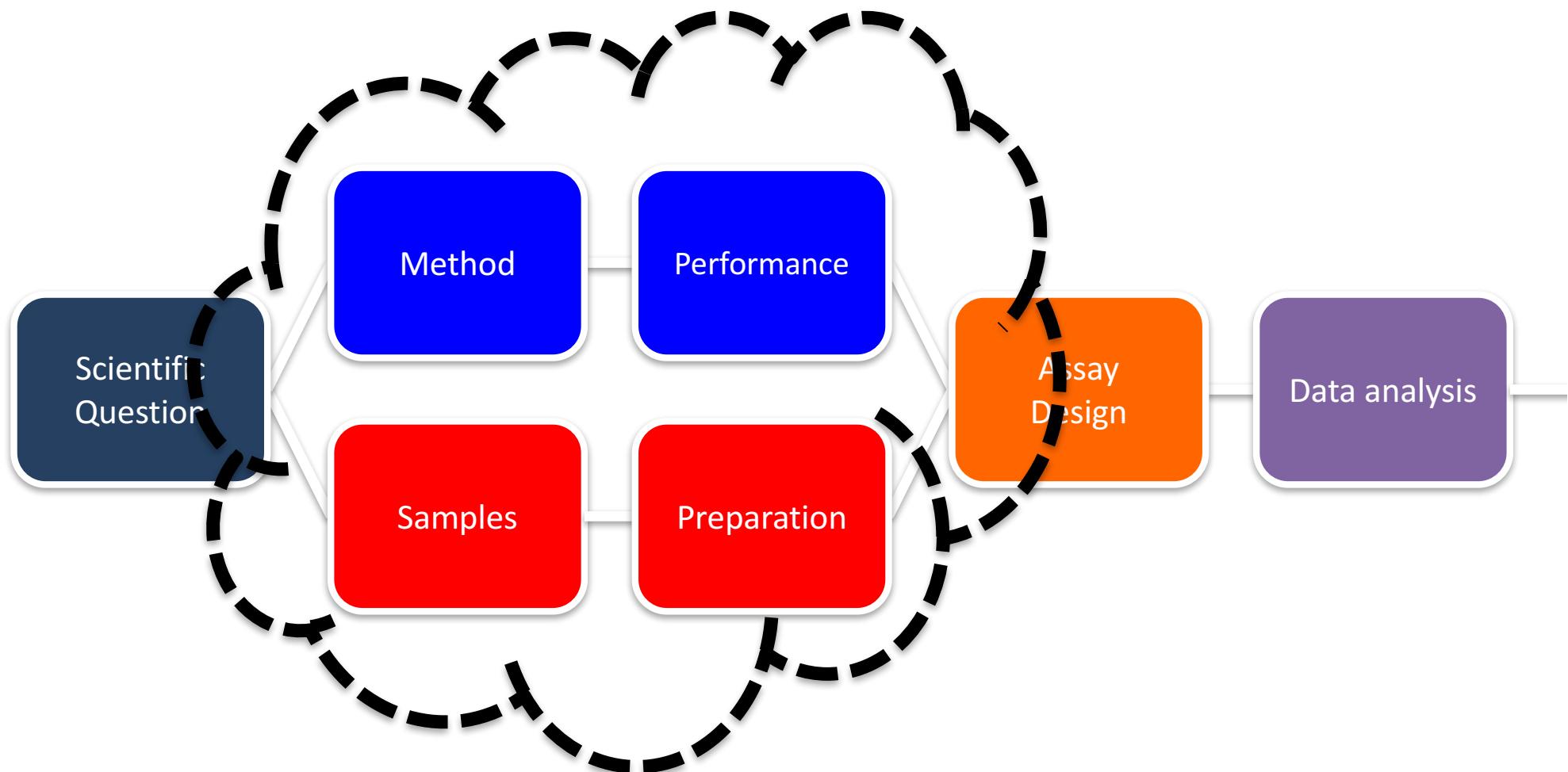


Fig. 1.11

|         |   |
|---------|---|
| 2DGE    | two-dimensional gel electrophoresis                   |
| AC      | affinity chromatography                               |
| CE      | capillary electrophoresis                             |
| CF      | chromatofocusing                                      |
| CGE     | capillary gel electrophoresis                         |
| CHAPS   | cholamidopropyltrimethyl-ammonio-propane-sulfonate    |
| DIGE    | difference gel electrophoresis                        |
| ESI     | electrospray ionization                               |
| ICAT    | isotope-coded affinity tags                           |
| IEC     | ion exchange chromatography                           |
| IEF     | isoelectric focusing                                  |
| IEX     | ion exchange chromatography                           |
| IMAC    | immobilized metal-affinity chromatography             |
| IPG     | immobilized pH gradient                               |
| LC      | liquid chromatography                                 |
| MALDI   | matrix assisted laser desorption ionization           |
| MS      | mass spectrometry                                     |
| NEPHGE  | nonequilibrium pH gradient electrophoresis            |
| PAGE    | polyacrylamide gel electrophoresis                    |
| pI      | isoelectric point                                     |
| PMF     | peptide mass fingerprinting                           |
| PSM     | peptide spectrum matches                              |
| PTM     | post-translational modifications                      |
| RP-HPLC | reversed-phase high performance liquid chromatography |
| SDS     | sodium dodecylsulfate                                 |
| SEC     | size exclusion chromatography                         |
| TOF     | time-of-flight  |

## Abbreviations

# EXPERIMENTAL PROTEOMICS



# SAMPLES AND PREPARATIONS

## Sample Types

Surgical tissues

Biopsy

Cells

Systemic Fluids

Proximal Fluids

## Sample Preparations

Tissue sectioning

Cell lysis

Extraction

Depletion

Chromatography

Digestion

Heat treatment

# EXAMPLES OF BODY FLUIDS

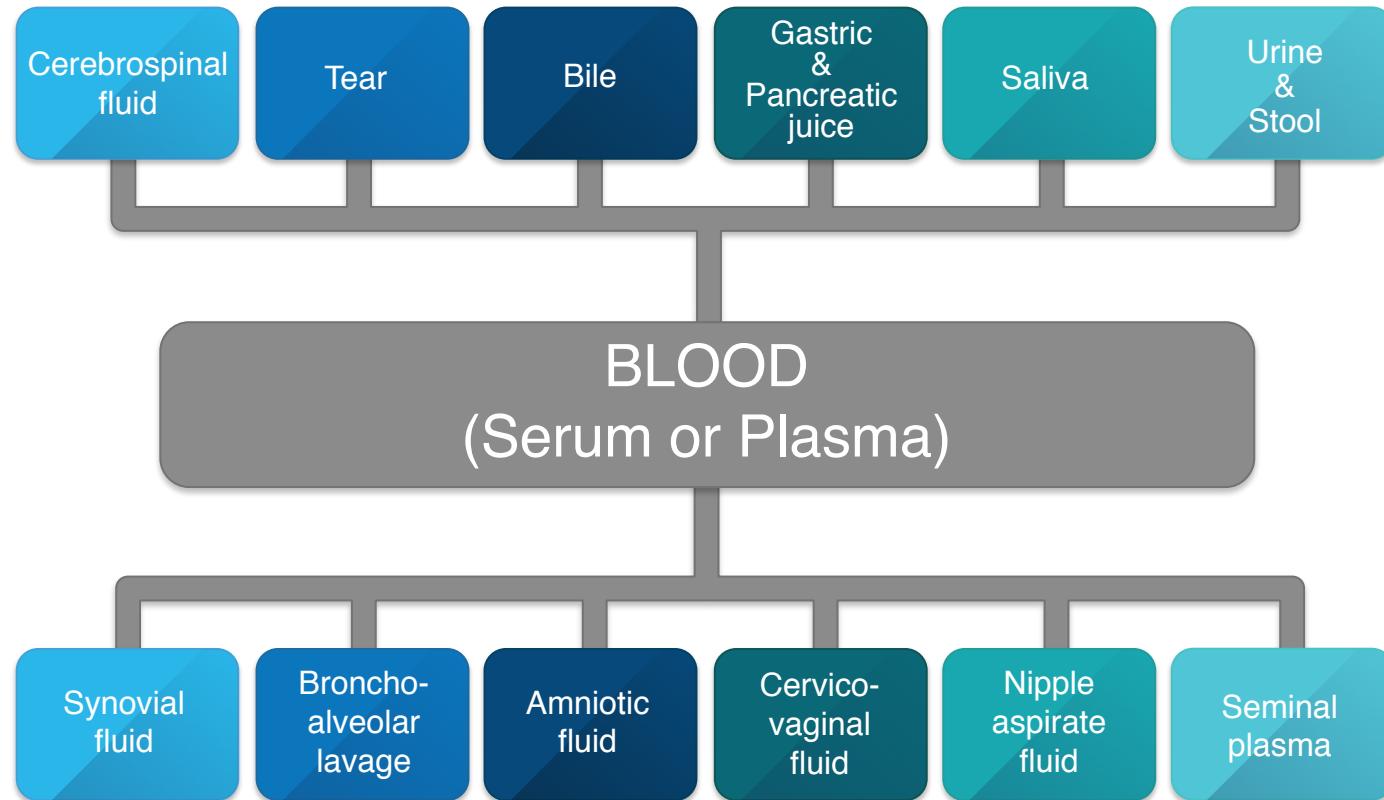


Illustration by Burcu Ayoglu

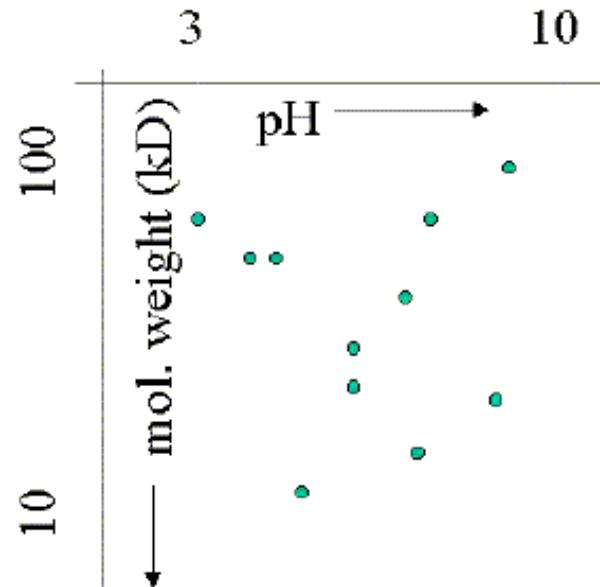
# WHY IS SAMPLE PREPARATION IMPORTANT?

- Structure and folding
- Function and activity
- Activation and modification status
- Interaction partners
- Localization
- Integrity
- Contamination

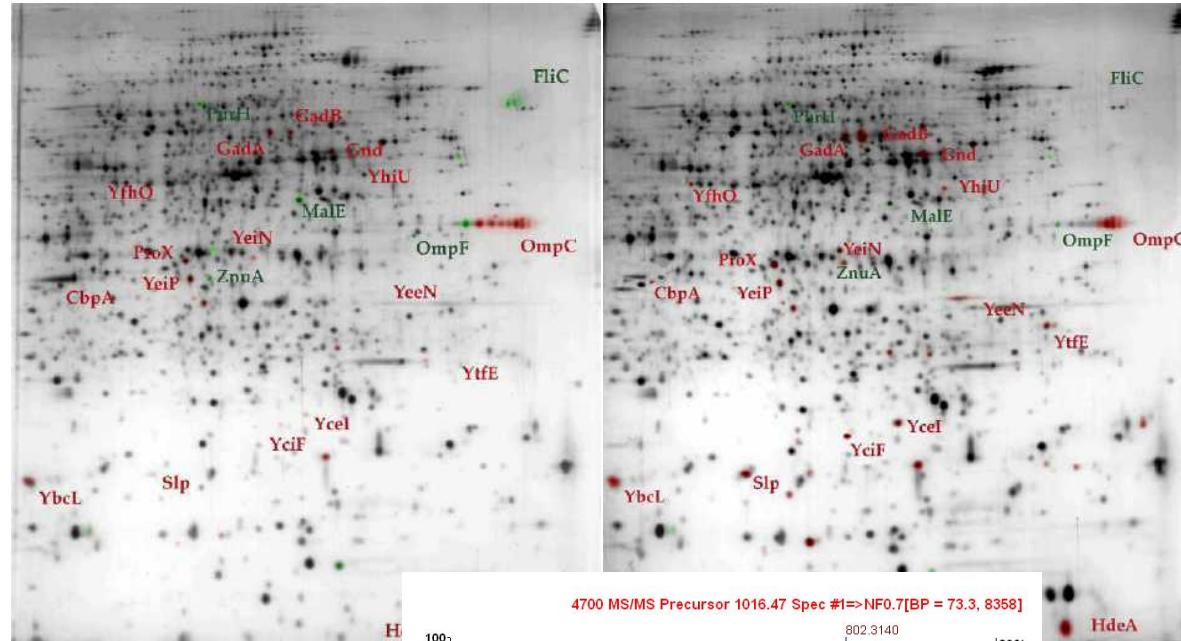
# **SEPARATION BASED PROTEOMICS**

# 2D-GEL ELECTROPHORESIS AND/OR MASS SPECTROMETRY ANALYSIS ARE THE CLASSICAL PROTEOMICS METHODS

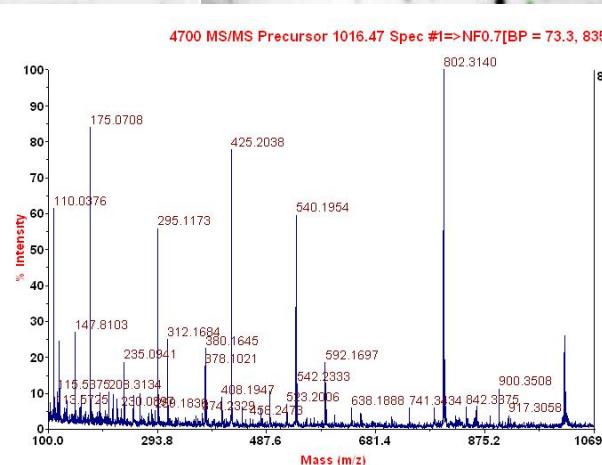
Separation of proteins in  
two dimensions



Identify differences between two samples

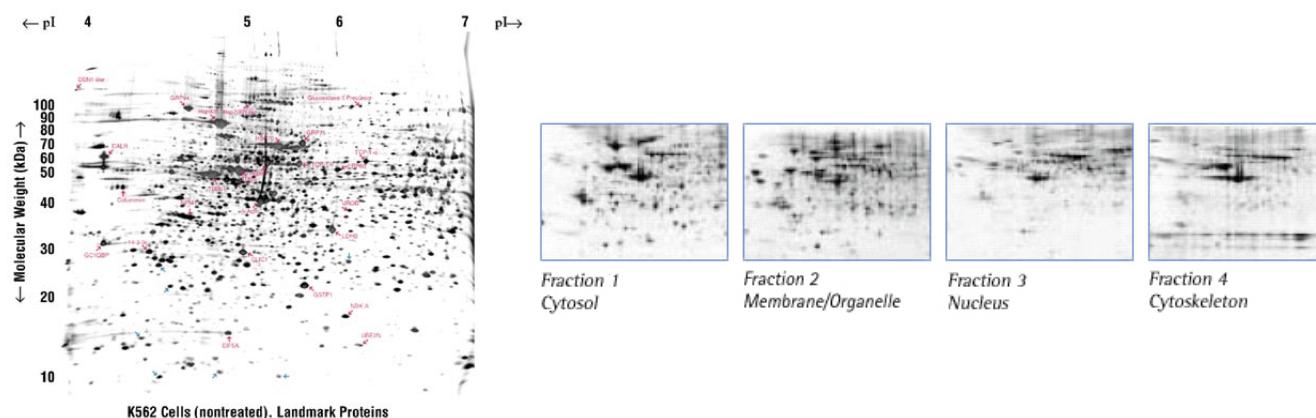


Identify proteins by mass spectrometry  
Based on peptide mass database searching

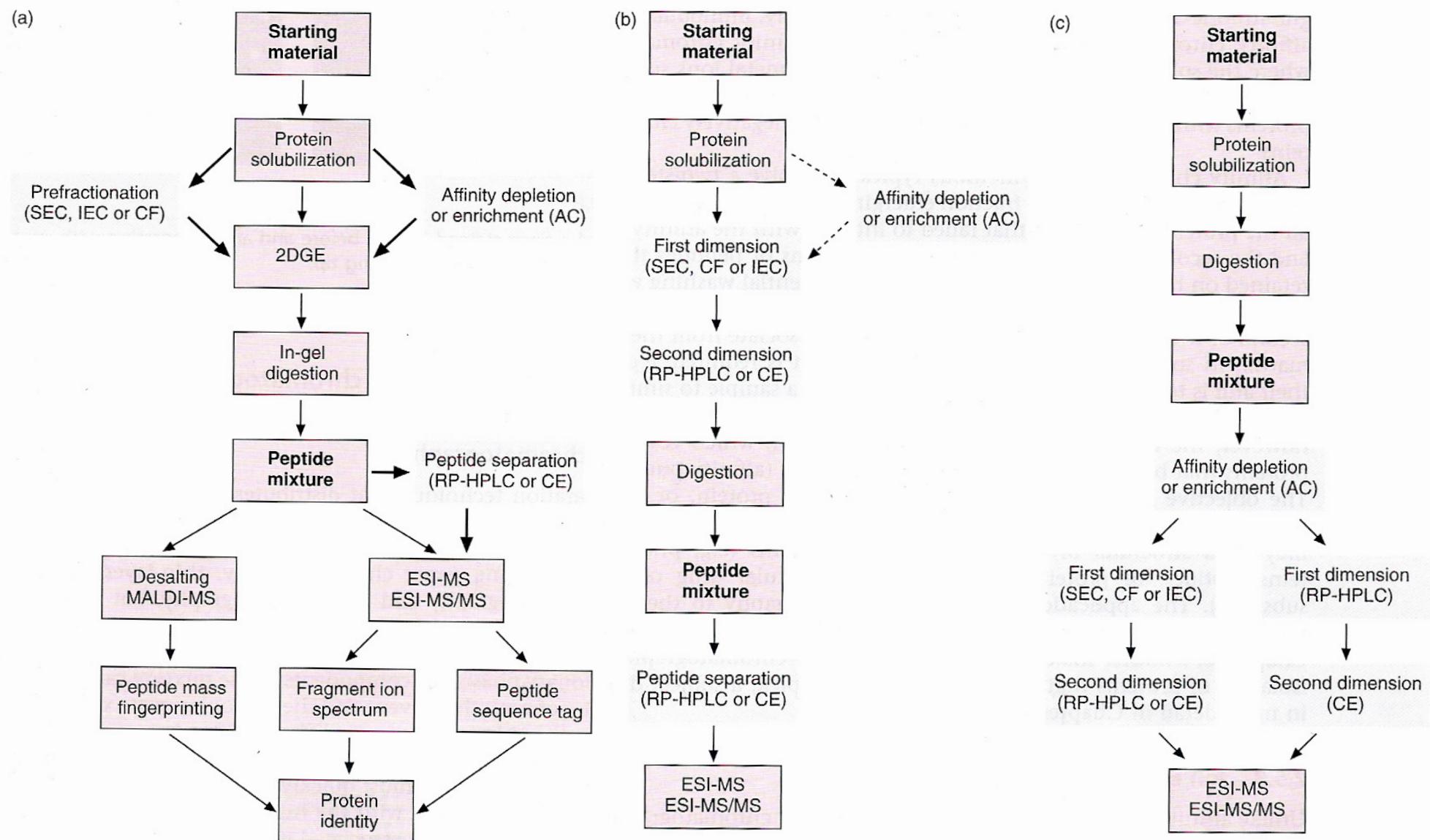


# LIMITATIONS OF 2DGE IN PROTEOMICS

- **Protein solubility:** Membrane proteins are under-represented on 2-D gels due to solubility problems
- **Proteomic coverage:** 2,000 proteins separated by 2-DE but many samples may contain >10,000 proteins
- **Dynamic range of protein abundance:**  $10^6$  for cells and tissue,  $10^{10}$  for body fluids
- **Sub-proteomes:** “Zoom”(narrow range) 2-D gels, Cellular, sub-cellular and protein fractionation
- **Representation:** The diversity among sample types and preparations
- **Automation:** Gel preparation cannot be automated compared to image analysis software followed isolation of individual spots picking for identification

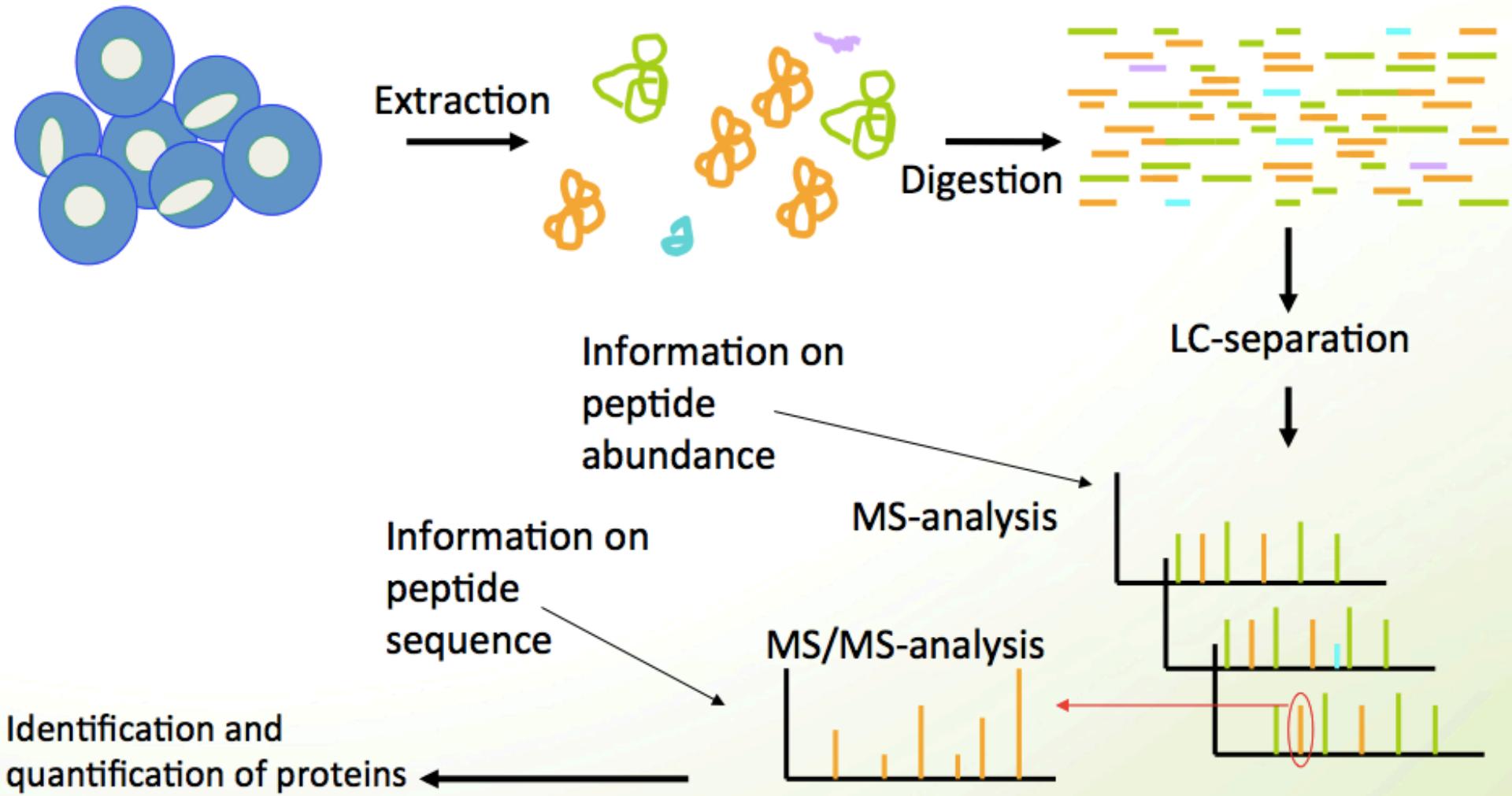


# LIQUID CHROMATOGRAPHY IN PROTEOMICS



# **MASS SPECTROMETRY ANALYSIS**

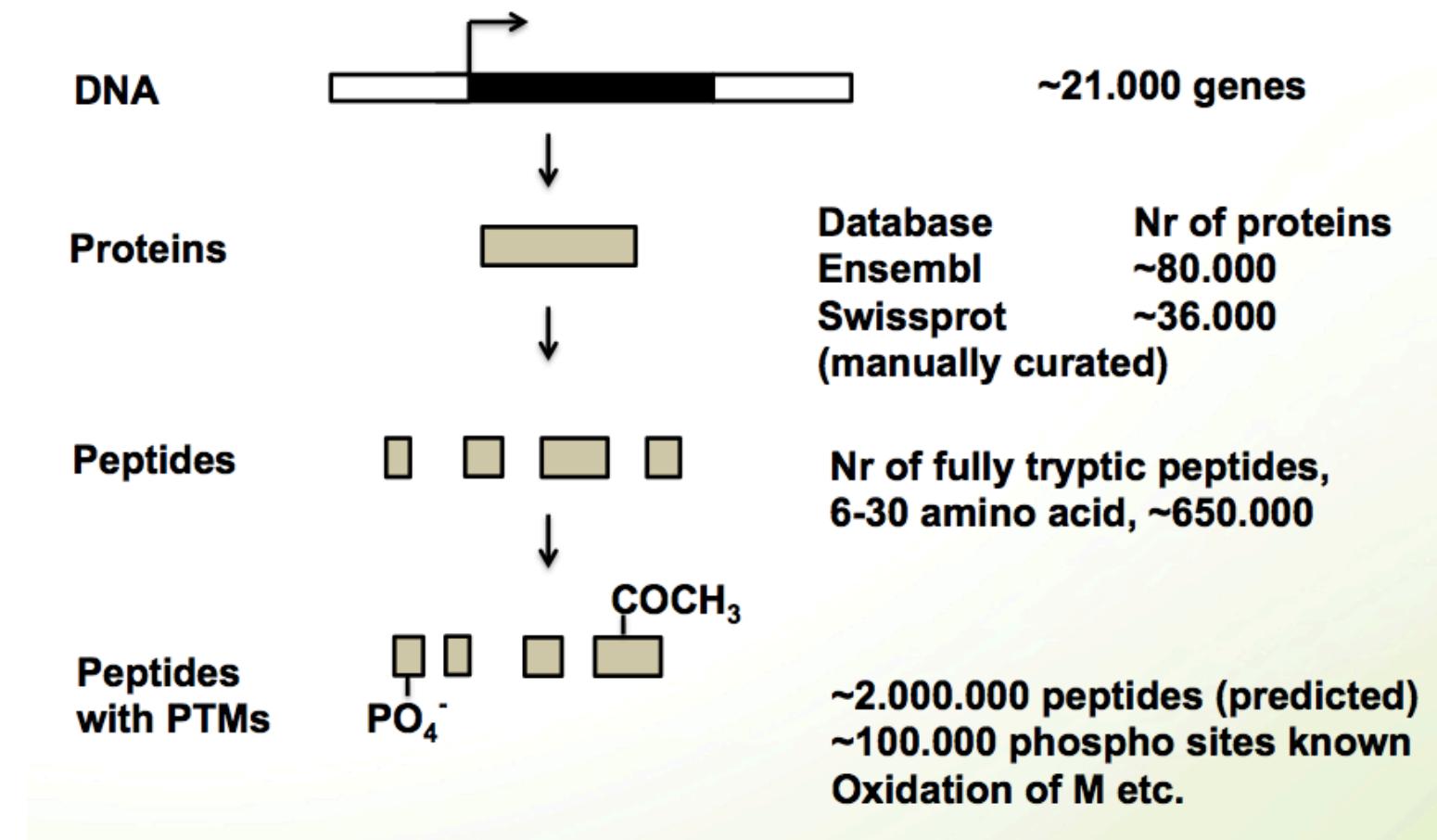
# What is MS-based proteomics



# WHY PEPTIDES INSTEAD OF PROTEINS?

- Easier to identify by MS
- Reverse phase Liquid chromatography (RP- LC) separates peptides better than proteins
- Most proteins give tryptic peptides that are soluble even if the protein itself is insoluble
- Peptides fragment give spectra to be sequenced.
- Peptides can be detected at lower levels

# WHY PRE-FRACTIONATE?



There is an upper mass detection limit  
Reduce sample complexity

# HOW TO PRE-FRACTIONATE?

TABLE 3.1 CLEAVAGE OF PROTEINS INTO PEPTIDES USING CHEMICAL AND ENZYMATIC REAGENTS

| Reagent                             | Cleavage properties                    |
|-------------------------------------|--|
| <i>Chemical agents</i>              |  |
| 70% formic acid                     | Asp-↓-Pro                              |
| Cyanogen bromide in 70% formic acid | Met-↓                                  |
| 2-Nitro-5-thiocyanobenzoate, pH 9   | ↓-Cys                                  |
| Hydroxylamine, pH 9                 | Asn-↓-Gly                              |
| Iodobenzoic acid in 50% acetic acid | Trp-↓                                  |
| <i>Endoprotease</i>                 |  |
| Trypsin                             | Arg/Lys-↓                              |
| Lys-C                               | Lys-↓                                  |
| Arg-C                               | Arg-↓                                  |
| Glu-C (bicarbonate)                 | Glu-↓                                  |
| Glu-C (phosphate)                   | Asp /Glu-↓                             |
| Asp-N                               | ↓-Asp                                  |
| Chymotrypsin                        | Phe/Tyr/Trp/Leu/Met-↓ (also Ile/Val-↓) |

The cleavage properties of all the endoproteases except Asp-N are dependent on the residue after the cleavage site not being proline.

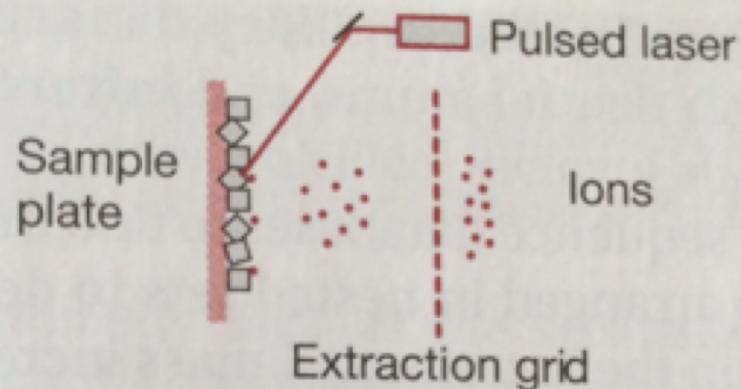
# SEPARATION OF PEPTIDES/PROTEINS

- Affinity chromatography
- Size exclusion chromatography
- Ion exchange chromatography
  - Anionic / cationic exchange
- Reverse-phase chromatography (HPLC)
  - hydrophobicity

# MS IONIZATION PRINCIPLES

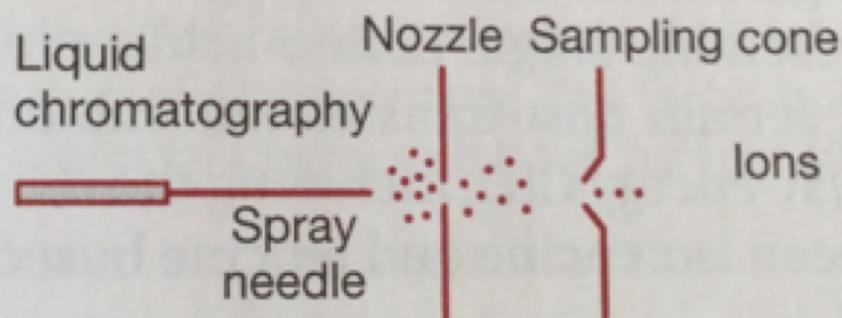
(a)

Matrix-assisted laser desorption/ionization (MALDI)



(b)

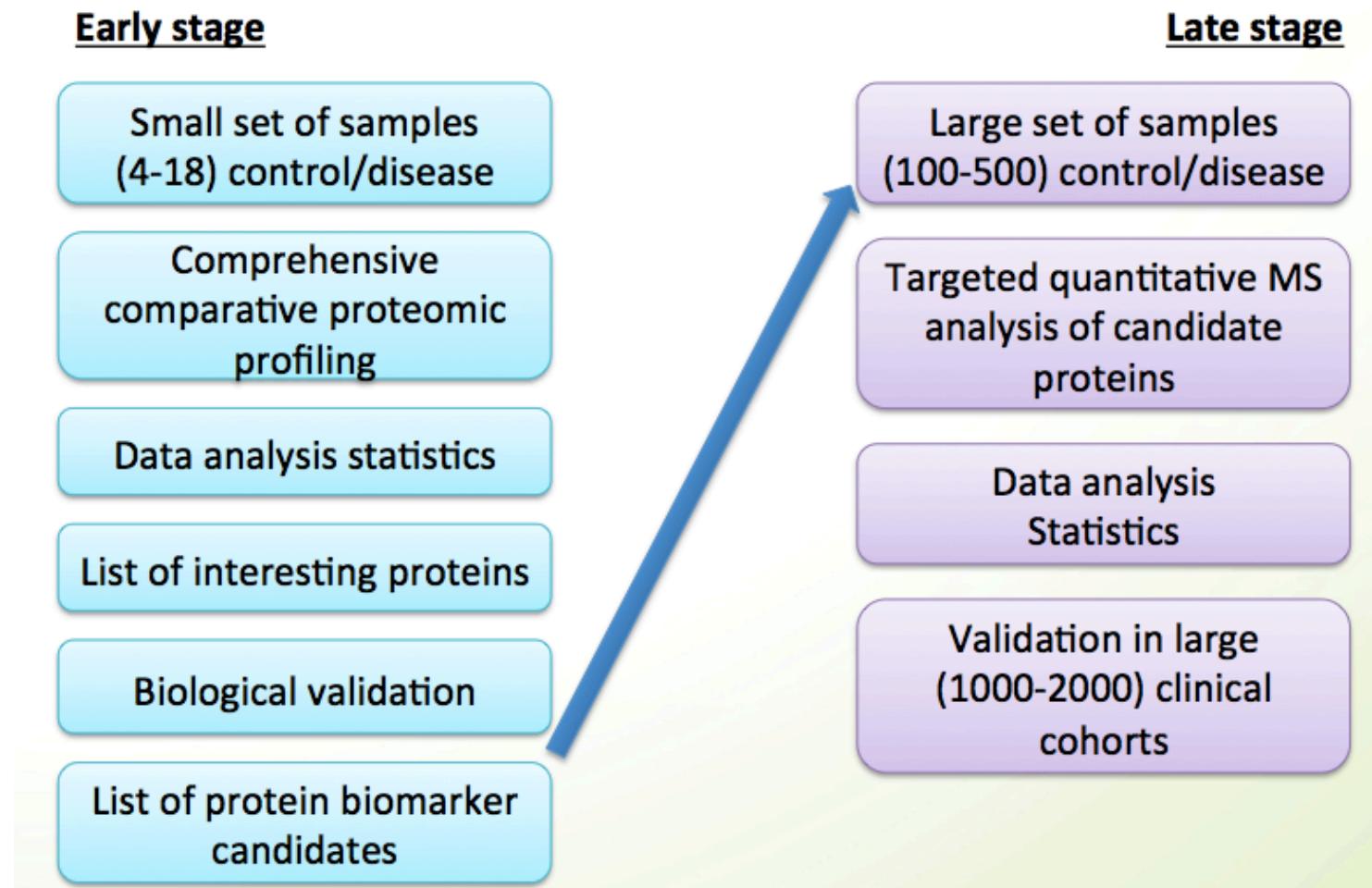
Electrospray ionization (ESI)



# WHAT TO IDENTIFY

| Aspartic acid | C <sub>4</sub> H <sub>7</sub> O <sub>2</sub> N              |   |     |           |
|---------------|---|---|-----|-----------|
| Cysteine      | C <sub>3</sub> H <sub>5</sub> ONS                           | C | Cys | 103.00919 |
| Glutamic acid | C <sub>5</sub> H <sub>7</sub> O <sub>3</sub> N              | E | Glu | 129.04259 |
| Glutamine     | C <sub>5</sub> H <sub>8</sub> O <sub>2</sub> N <sub>2</sub> | Q | Gln | 128.05858 |
| Glycine       | C <sub>2</sub> H <sub>3</sub> ON                            | G | Gly | 57.02146  |
| Histidine     | C <sub>6</sub> H <sub>7</sub> ON <sub>3</sub>               | H | His | 137.05891 |
| Isoleucine    | C <sub>6</sub> H <sub>11</sub> ON                           | I | Ile | 113.08406 |
| Leucine       | C <sub>6</sub> H <sub>11</sub> ON                           | L | Leu | 113.08406 |
| Lysine        | C <sub>6</sub> H <sub>12</sub> ON <sub>2</sub>              | K | Lys | 128.09496 |
| Methionine    | C <sub>5</sub> H <sub>9</sub> ONS                           | M | Met | 131.04049 |
| Phenylalanine | C <sub>9</sub> H <sub>9</sub> ON                            | F | Phe | 147.06841 |

# “CLASSICAL” MS BASED BIOMARKER DISCOVERY



# MS BASED QUANTIFICATION

- Label free quantification
- Stable isotopic labeling

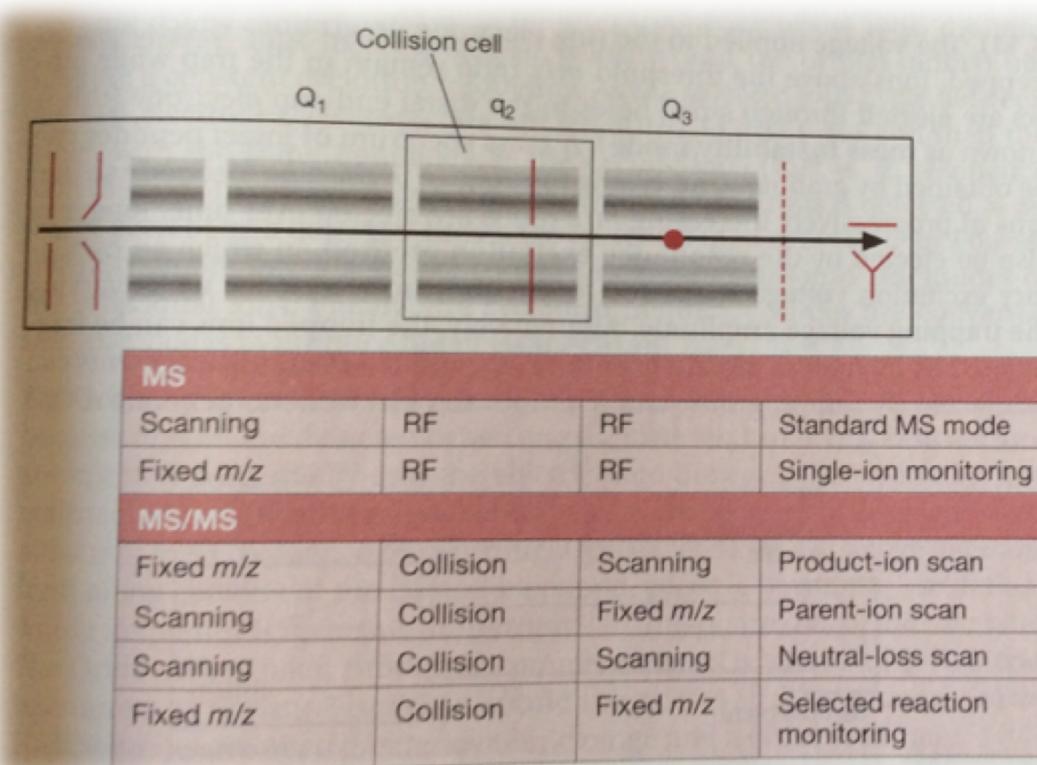
## *Chemical labeling*

- ICAT
- iTRAQ (4-plex, 8-plex)
- TMT (6-plex, 10-plex)

## *Metabolic labeling*

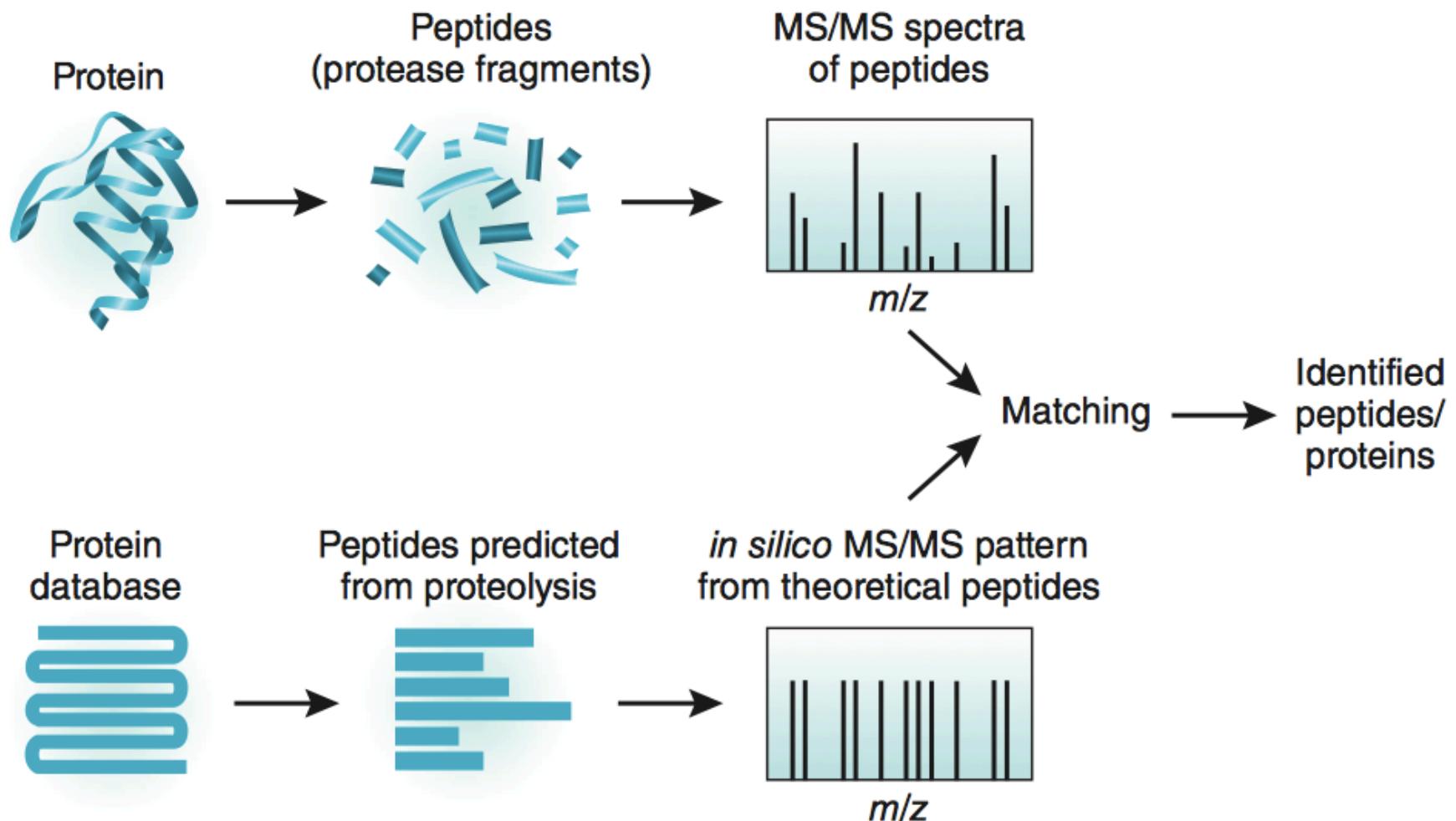
- SILAC or Super-SILAC
- N15

# PEPTIDE SELECTOR



QQQ

# DATA ANALYSIS IN MS



# Quantitative proteomics with mass spectrometry

more reproducible and sensitive than gel-based methods

## ICAT: isotope-coded affinity tags

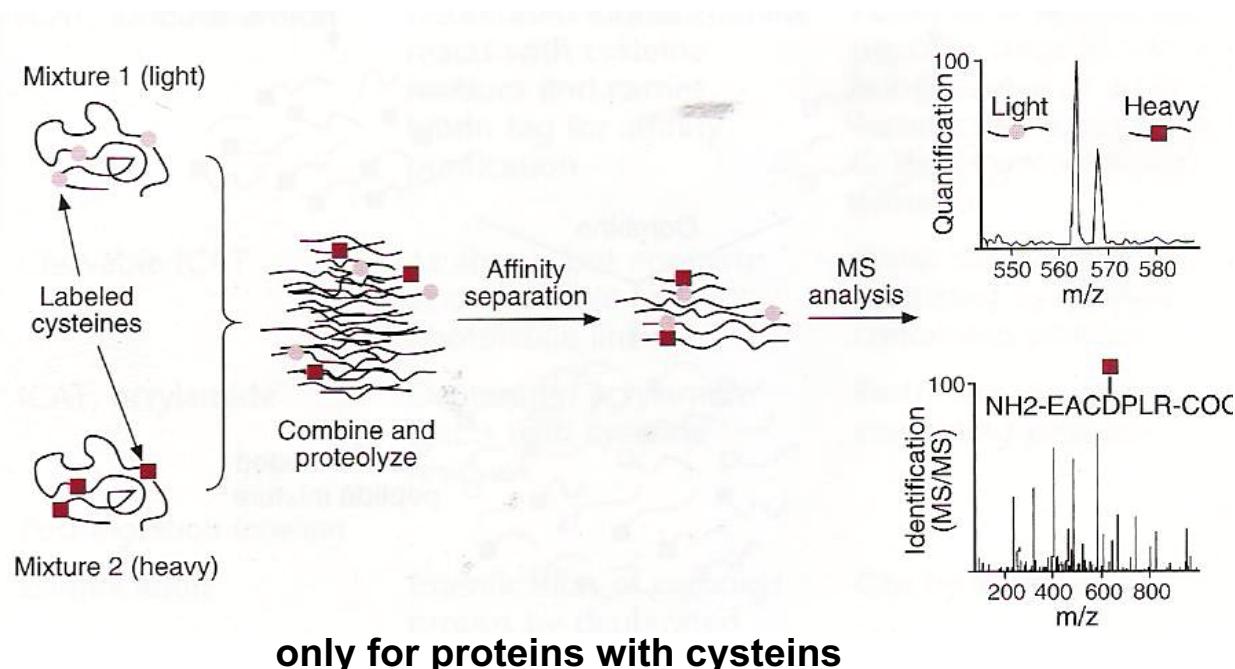


Fig 4.4

## Isotope tagging *in vivo*

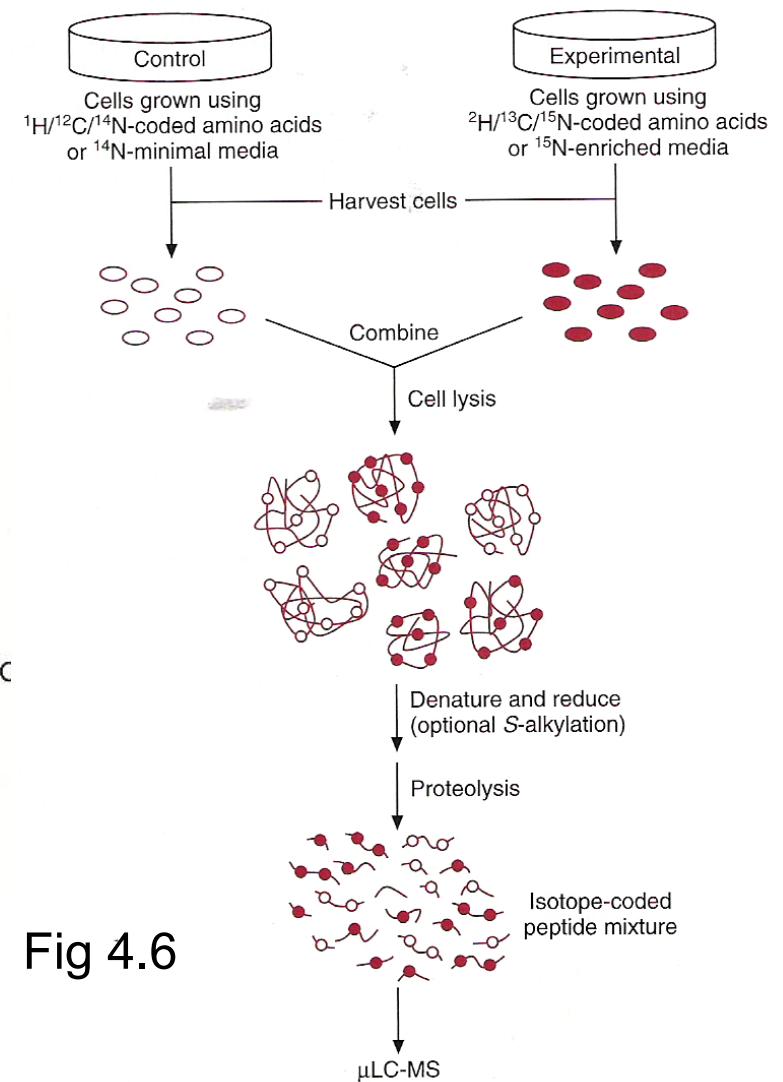


Fig 4.6

# SOME LIMITATIONS OF MS

- Reproducibility
- Low through-put (number of samples)
- Limited dynamic range (2-3)
- Limited sensitivity
- Data output dependent on software algorithm
- Serum/Plasma are not preferred as sample

# SOME ADVANTAGES OF MS

- Protein (sequence) identification
- Analysis of post-translation modifications
- Discovery-driven or targeted analysis
- (Absolute) Protein quantification
- High-throughput (proteins per sample)
- Growing data bases
- Widely used

# MS BASED PROTEOMES

## ARTICLE

doi:10.1038/nature13319

### Mass-spectrometry-based draft of the human proteome

Mathias Wilhelm<sup>1,2\*</sup>, Judith Schlegl<sup>2\*</sup>, Hannes Hahne<sup>1\*</sup>, Amin Moghaddas Gholami<sup>1\*</sup>, Marcus Lieberenz<sup>2</sup>, Mikhail M. Savitski<sup>3</sup>, Emanuel Ziegler<sup>2</sup>, Lars Butzmann<sup>1</sup>, Siegfried Gessulat<sup>2</sup>, Harald Marx<sup>1</sup>, Toby Mathieson<sup>3</sup>, Simone Lemer<sup>1</sup>, Karsten Schnatbaum<sup>4</sup>, Ulf Reimer<sup>4</sup>, Holger Wenschuh<sup>4</sup>, Martin Mollenhauer<sup>5</sup>, Julia Slotta-Huspenina<sup>5</sup>, Joos-Hendrik Boese<sup>5</sup>, Marcus Bantscheff<sup>3</sup>, Anja Gerstmair<sup>2</sup>, Franz Faerber<sup>2</sup> & Bernhard Kuster<sup>1,6</sup>

Proteomes are characterized by large protein-abundance differences, cell-type- and time-dependent expression patterns and post-translational modifications, all of which carry biological information that is not accessible by genomics or transcriptomics. Here we present a mass-spectrometry-based draft of the human proteome and a public, high-performance, in-memory database for real-time analysis of terabytes of big data, called ProteomicsDB. The information assembled from human tissues, cell lines and body fluids enabled estimation of the size of the protein-coding genome, and identified organ-specific proteins and a large number of translated lincRNAs (long intergenic non-coding RNAs). Analysis of messenger RNA and protein-expression profiles of human tissues revealed conserved control of protein abundance, and integration of drug-sensitivity data enabled the identification of proteins predicting resistance or sensitivity. The proteome profiles also hold considerable promise for analysing the composition and stoichiometry of protein complexes. ProteomicsDB thus enables navigation of proteomes, provides biological insight and fosters the development of proteomic technology.

## ARTICLE

doi:10.1038/nature13302

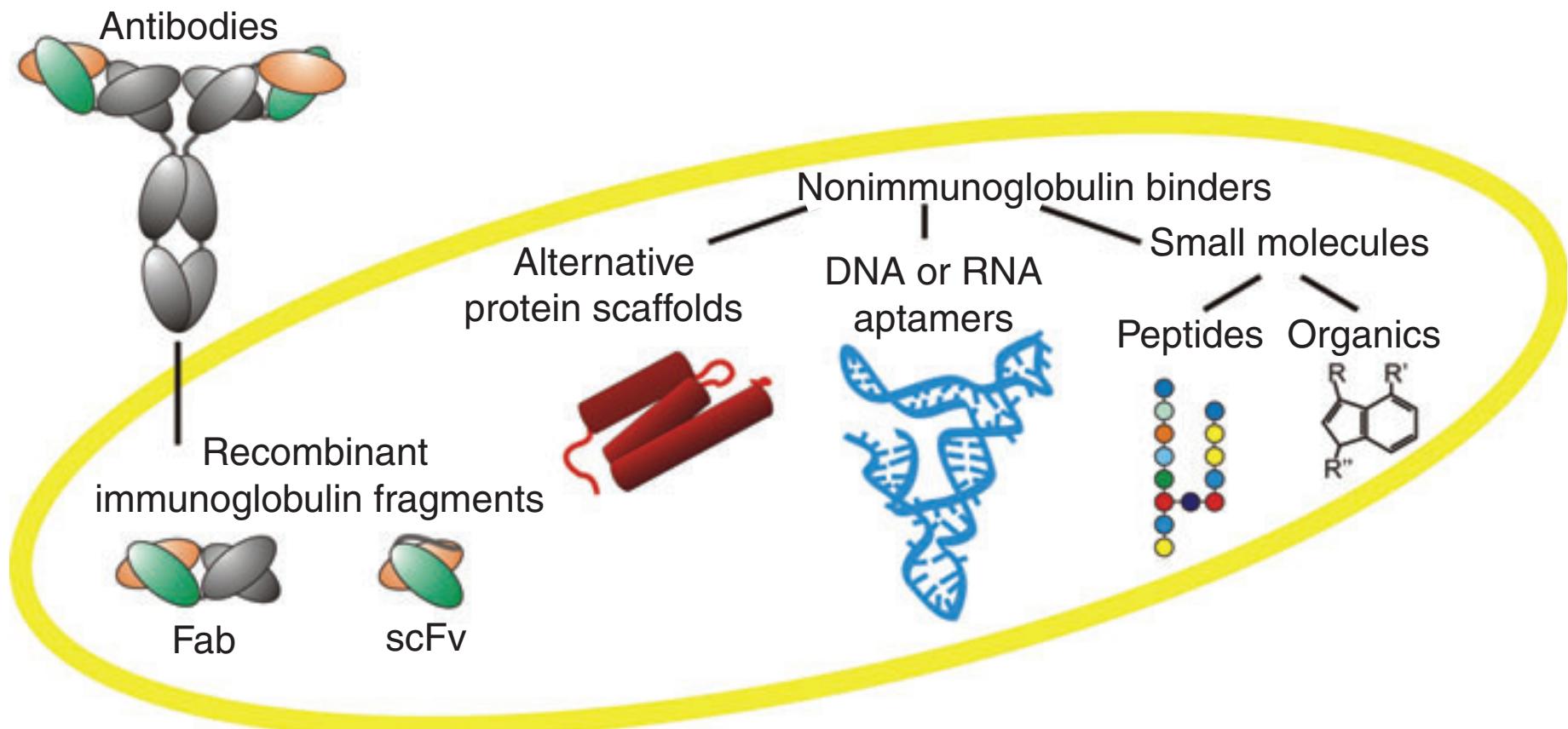
### A draft map of the human proteome

Min-Sik Kim<sup>1,2</sup>, Sneha M. Pinto<sup>3</sup>, Derese Getnet<sup>1,4</sup>, Raja Sekhar Nirujogi<sup>3</sup>, Srikanth S. Manda<sup>3</sup>, Raghothama Chaerkady<sup>1,2</sup>, Anil K. Madugundu<sup>3</sup>, Dhanashree S. Kelkar<sup>3</sup>, Ruth Isserlin<sup>5</sup>, Shobhit Jain<sup>1</sup>, Joji K. Thomas<sup>3</sup>, Babylakshmi Muthusamy<sup>3</sup>, Pamela Leal-Rojas<sup>4</sup>, Praveen Kumar<sup>6</sup>, Nandini A. Sahasrabudhe<sup>6</sup>, Lavanya Balakrishnan<sup>7</sup>, Jayshree Advani<sup>7</sup>, Bijesh George<sup>8</sup>, Santosh Renuse<sup>3</sup>, Lakshmi Dhevvi N. Selvan<sup>3</sup>, Arun H. Patil<sup>3</sup>, Vishalakshi Nanjappa<sup>3</sup>, Aneesa Radhakrishnan<sup>3</sup>, Samarjeet Prasad<sup>1</sup>, Tejaswini Subbannayya<sup>3</sup>, Rajesh Raju<sup>3</sup>, Manish Kumar<sup>3</sup>, Sreelakshmi K. Sreenivasamurthy<sup>3</sup>, Arivusudar Marimuthu<sup>3</sup>, Gajanan J. Sathe<sup>3</sup>, Sandip Chavan<sup>3</sup>, Keshava K. Datta<sup>3</sup>, Yashwanth Subbannayya<sup>3</sup>, Apeksha Sahu<sup>3</sup>, Soujanya D. Yelamanchi<sup>3</sup>, Savita Jayaram<sup>3</sup>, Pavithra Rajagopalan<sup>3</sup>, Jyoti Sharma<sup>4</sup>, Krishna R. Murthy<sup>4</sup>, Nazia Syed<sup>3</sup>, Renu Goel<sup>3</sup>, Afaque A. Khan<sup>3</sup>, Sartaj Ahmad<sup>3</sup>, Gourav Dey<sup>3</sup>, Keshav Mudgal<sup>7</sup>, Aditi Chatterjee<sup>3</sup>, Tai-Chung Huang<sup>1</sup>, Jun Zhong<sup>1</sup>, Xinyan Wu<sup>1,2</sup>, Patrick G. Shaw<sup>1</sup>, Donald Freed<sup>1</sup>, Muhammad S. Zahari<sup>2</sup>, Kanchan K. Mukherjee<sup>8</sup>, Subramanian Shankar<sup>9</sup>, Anita Mahadevan<sup>10,11</sup>, Henry Lam<sup>12</sup>, Christopher J. Mitchell<sup>11</sup>, Susarla Krishna Shankar<sup>10,11</sup>, Partha Sarathy Satishchandra<sup>13</sup>, John T. Schroeder<sup>14</sup>, Ravi Sirdeshmukh<sup>3</sup>, Anirban Maitra<sup>15,16</sup>, Steven D. Leach<sup>1,17</sup>, Charles G. Drake<sup>16,18</sup>, Marc K. Halushka<sup>15</sup>, T. S. Keshava Prasad<sup>3</sup>, Ralph H. Hruban<sup>15,16</sup>, Candace L. Kerr<sup>19†</sup>, Gary D. Bader<sup>5</sup>, Christine A. Iacobuzio-Donahue<sup>15,16,17</sup>, Harsha Gowda<sup>3</sup> & Akhilesh Pandey<sup>1,2,3,4,15,16,20</sup>

The availability of human genome sequence has transformed biomedical research over the past decade. However, an equivalent map for the human proteome with direct measurements of proteins and peptides does not exist yet. Here we present a draft map of the human proteome using high-resolution Fourier-transform mass spectrometry. In-depth proteomic profiling of 30 histologically normal human samples, including 17 adult tissues, 7 fetal tissues and 6 purified primary haematopoietic cells, resulted in identification of proteins encoded by 17,294 genes accounting for approximately 84% of the total annotated protein-coding genes in humans. A unique and comprehensive strategy for proteogenomic analysis enabled us to discover a number of novel protein-coding regions, which includes translated pseudogenes, non-coding RNAs and upstream open reading frames. This large human proteome catalogue (available as an interactive web-based resource at <http://www.humanproteomap.org>) will complement available human genome and transcriptome data to accelerate biomedical research in health and disease.

# **EXAMPLES OF ANTIBODY-BASED ANALYSIS METHODS**

# AFFINITY BINDERS



Taken from – Taussig et al (2007) Nat Methods

# TISSUE-BASED MAP OF THE HUMAN PROTEOME

RESEARCH

**RESEARCH ARTICLE SUMMARY**

PROTEOMICS

# Tissue-based map of the human proteome

Mathias Ahlbom,<sup>1</sup> Linn Fagerberg,<sup>2</sup> Björn Björk,<sup>3</sup> Evelina Lindskog,<sup>4</sup> Per Olsvärd,<sup>5</sup> Adil Mardinoglu,<sup>6</sup> Åsa Sverstorp,<sup>7</sup> Caroline Kampf,<sup>8</sup> Evelina Sjöstedt,<sup>9</sup> Anna Asplund,<sup>10</sup> IngMarie Olsson,<sup>11</sup> Karolina Edlund,<sup>12</sup> Emma Lundberg,<sup>13</sup> Sanjay Navani,<sup>14</sup> Cristina Al-Khalili Szczerba,<sup>15</sup> Jacob Odeberg,<sup>16</sup> Dijana Djericinovic,<sup>17</sup> Jenny Ottosson Takken,<sup>18</sup> Sophia Hober,<sup>19</sup> Tove Alm,<sup>20</sup> Per-Henrik Edqvist,<sup>21</sup> Holger Berling,<sup>22</sup> Hanna Tegel,<sup>23</sup> Jan Mulder,<sup>24</sup> Johan Rockberg,<sup>25</sup> Peter Nilsson,<sup>26</sup> Jochen M. Schwenk,<sup>27</sup> Maria Hamsten,<sup>28</sup> Karl von Felilitzen,<sup>29</sup> Mattias Forsberg,<sup>30</sup> Lukas Persson,<sup>31</sup> Fredric Johansson,<sup>32</sup> Martin Zwahlen,<sup>33</sup> Paul von Heijne,<sup>34</sup> Jens Nielsen,<sup>35</sup> Fredrik Pönöniemi,<sup>36</sup> and Mikael Lindström,<sup>37</sup> on behalf of the Swedish COVID-19 Research Consortium.

**INTRODUCTION:** Resolving the molecular details of proteome variation in the different tissues and organs of the human body would greatly increase our knowledge of human biology and disease. Here, we present a map of the human tissue proteome based on quantitative transcriptomics on a tissue and organ level combined with protein profiling using surface plasmon resonance (SPR) and mass spectrometry. We describe the spatial localization of proteins down to the spatial resolution. We compare *a* priori proteome analysis with post-translational modifications. Furthermore, this is enrichment analysis of the proteome. In addition, we compare the proteome with the transcriptome. The results show that almost half of the genes expressed in all analyzed tissues, which suggests that certain gene products are needed in all cells to maintain "housekeeping" functions, such as cytoskeletal components.

global analysis of gene expression profiles of the secreted and membrane proteins, as well as an analysis of the expression profiles for all proteins targeted by pharmaceutical drugs and proteins implicated in cancer.

**RATIONALE:** We have used an integrative omics approach to study the spatial human proteome. Samples representing all major tissues and organs ( $n = 44$ ) in the human body have been analyzed based on 24,928 antibodies corresponding to 45,975 protein-coding genes, complemented with RNA-seq sequencing data for 32 of the tissues. The antibodies have been used to produce more than 13 million test results. The figure shows a schematic diagram of a human figure with various tissues highlighted in pink, corresponding to the antibody distribution data. A legend on the right side indicates the color coding for different antibody properties: yellow for soluble, red for membrane, and green for secreted.

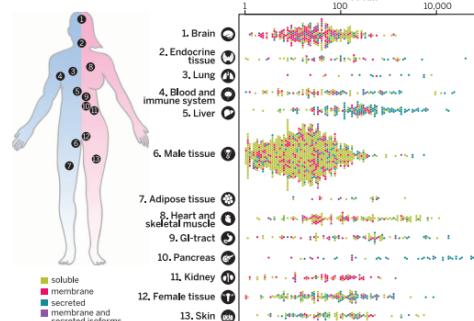
| Tissue Type                  | Antibody Properties |
|------------------------------|---------------------|
| 2. Endocrine tissue          | Soluble             |
| 3. Lung                      | Membrane            |
| 4. Blood and immune system   | Soluble, Secreted   |
| 5. Liver                     | Secreted            |
| 6. Male tissue               | Soluble, Secreted   |
| 7. Adipose tissue            | Soluble             |
| 8. Heart and skeletal muscle | Membrane            |
| 9. GI-tract                  | Secreted            |
| 10. Pancreas                 | Soluble, Secreted   |
| 11. Kidney                   | Secreted            |
| 12. Female tissue            | Soluble, Secreted   |

**RESULTS:** We report a genome-wide analysis of the tissue specificity of RNA and protein expression covering more than 90% of the nutant protein-coding genes in the mouse genome. The results show that the distribution of tissue-specific gene expression is highly correlated with the distribution of membrane and secreted isoforms (Fig. 1). In contrast, the distribution of membrane and secreted isoforms is not correlated with the distribution of tissue-specific gene expression.

f  
e  
d  
n  
x  
s  
al  
ne  
n

e.

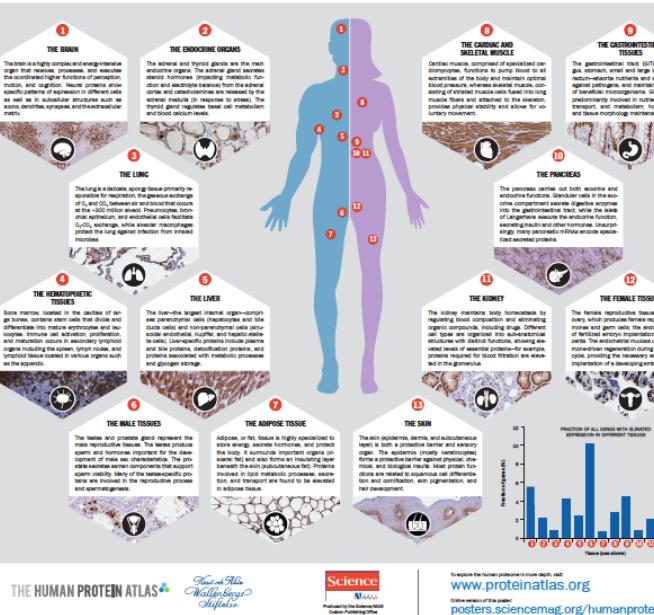
+



Jhlen et al (2015)  
"Tissue-based map of the human proteome"  
Science 347: 394 (January 23)

# THE HUMAN PROTEOME

**The Power of Proteins.** If DNA can be equated with the blueprint for a home, then proteins provide the bricks and mortar, plumbing, paint—essentially everything that makes up the house. The human genome consists of approximately 20,000 protein-encoding genes. This poster summarizes the multiple on-going antibody-based proteomic-based proteome projects and where in the human body this research is focused. For more detailed information, view [www.proteinatia.org](http://www.proteinatia.org)



THE SECRETOMI  
MEMBRANE PR

Both secreted and membrane proteins play crucial roles in many physiological processes. Membrane proteins include cytokines, growth factors, among others. Secreted proteins include ion channel apertors, enzymes, receptors, and other proteins. Approximately 10% are predicted to interact with another 5,500 extracellular proteins.

THE ISOFORM P

The existence of a vast number of proteins in each cell endows the man proteome with both structural and functional diversity. The forms are produced by alternative splicing, post-translational modifications, protein-protein interaction, protein recombination, and gene conversion. Variations in gene expression and post-translational modifications also result from mutations in protein-coding regions. The proportion of proteins with different variants yields the yield of protein isoforms. The almost limitless number of modifications and additional variants produces a diverse proteome.

THE CANCER PR

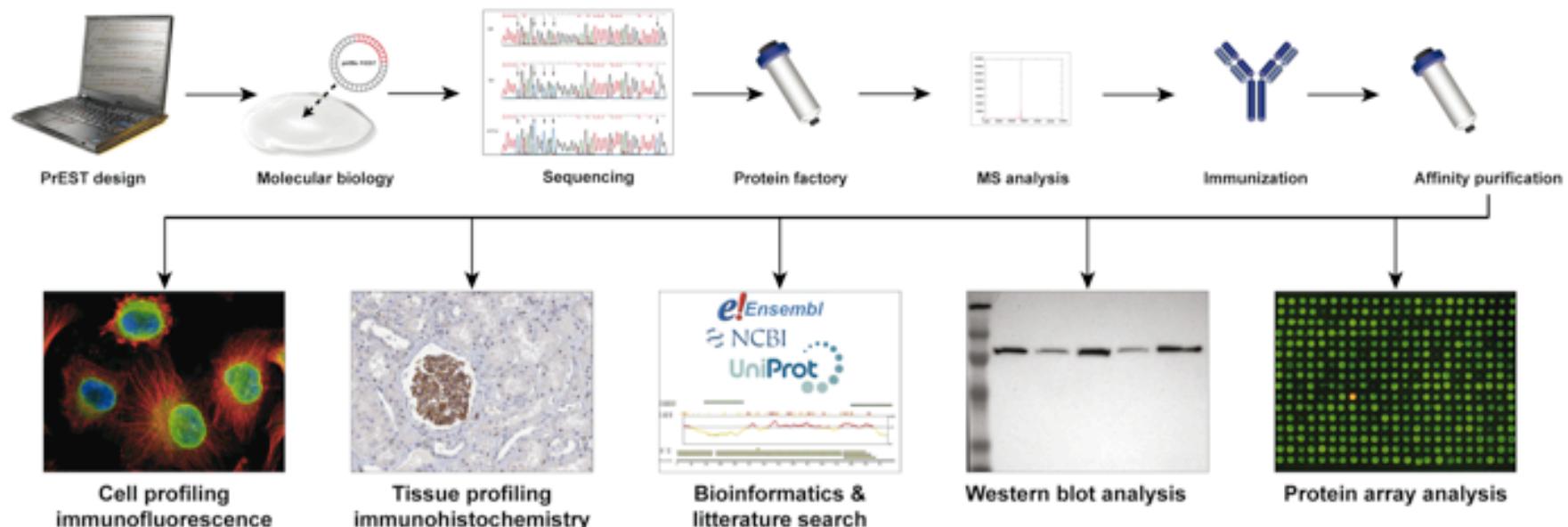
Over 500 genes have been transformed, and genes are essential for function. However, expression, or expression can contribute to dysfunction. Dysregulated expression, structural rearrangement, specific gene transcription through mechanisms. Furthermore, small insertions or deletion of function in the

THE DRUGGABLE

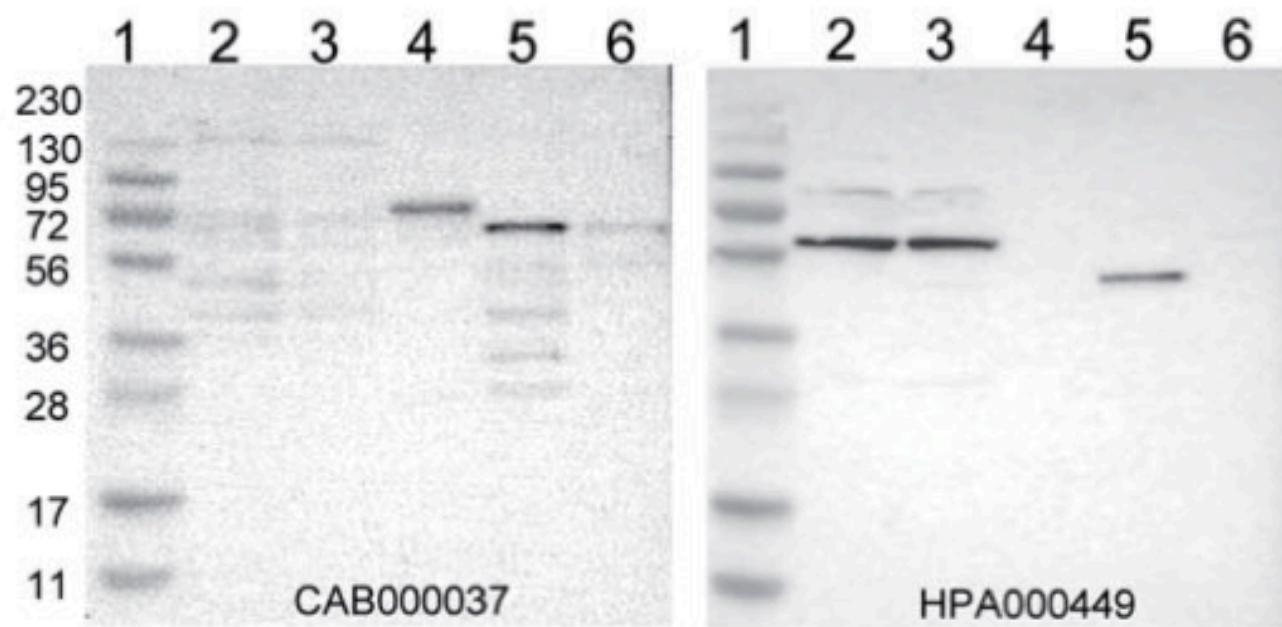
Most pharmaceuticals and modulating agents belong to four main families: ion channels, enzymes, and Drug Admetabolites, acting approximately most on signal transduction, converting extracellular messengers. Antibody-based molecules penetrate the plasma membrane to target cell surface proteins, while small molecules can both intracellular and extracellular.

Poster in Science (November 7, 2015)  
New interactive database ([www.proteinatlas.org](http://www.proteinatlas.org))

# ANTIBODY GENERATION AND APPLICATION

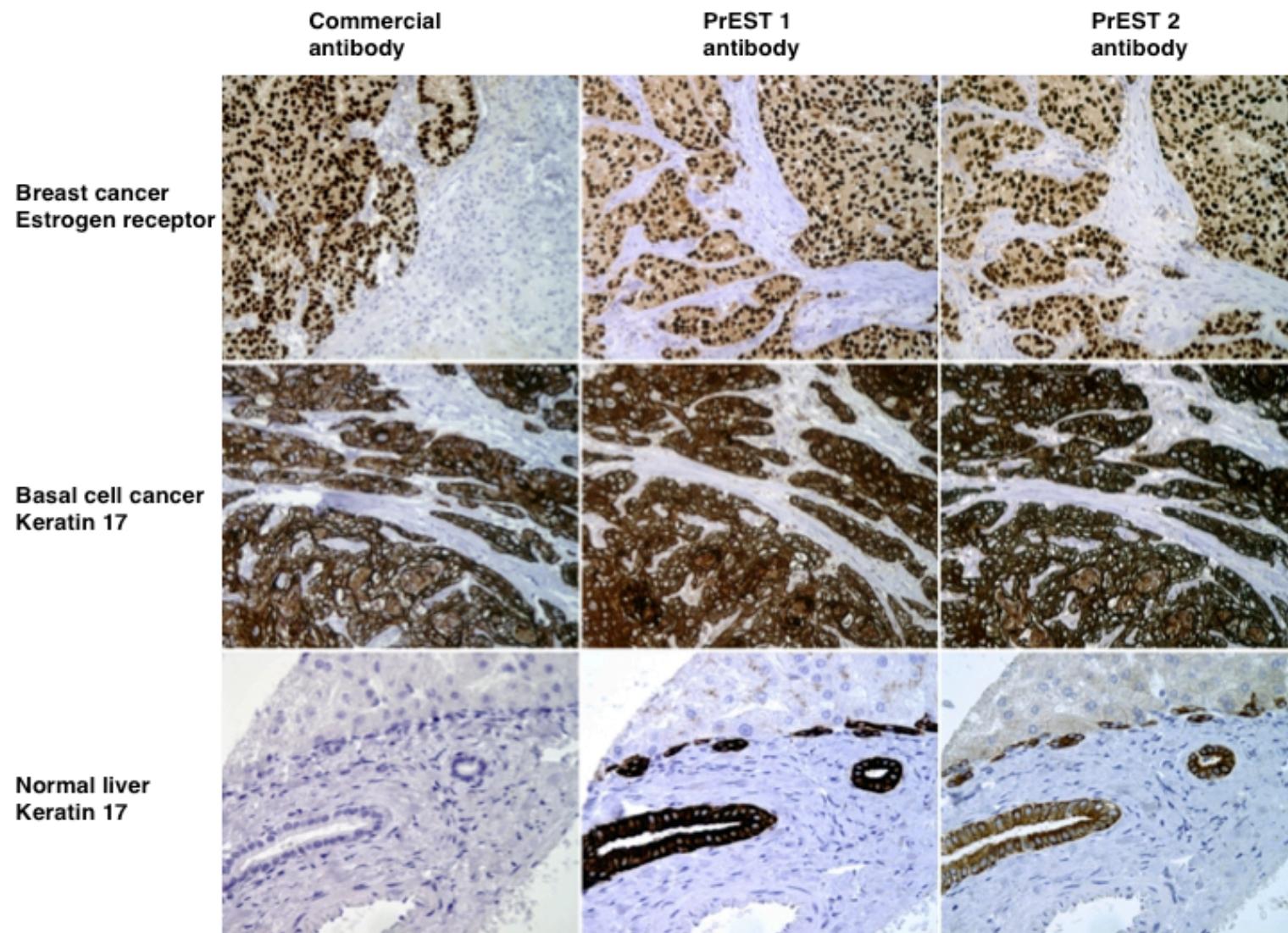


# WESTERN BLOTH ANALYSIS



Algenäs et al (2014) *Biotechnol J.*

# IMMUNOHISTOCHEMISTRY SPATIAL RESOLUTION OF PROTEIN EXPRESSION



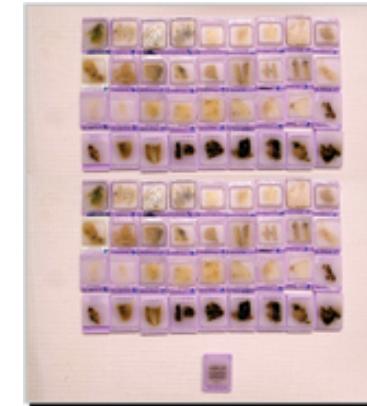
# Tissue Microarrays (TMA)



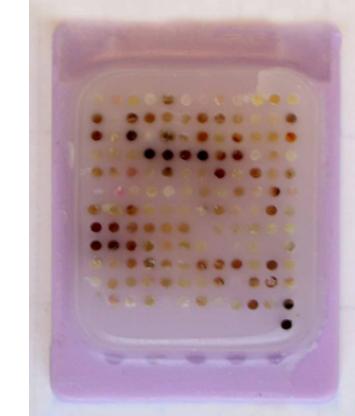
Collection of tissues



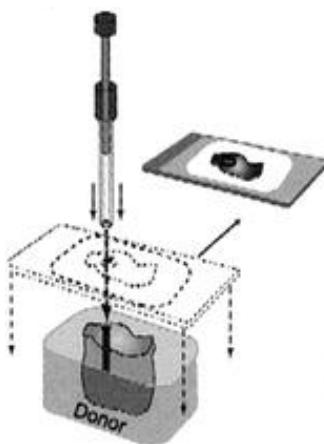
Annotation



TMA design

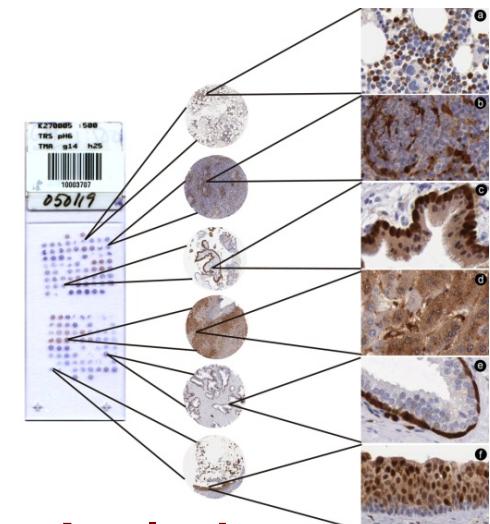


TMA block



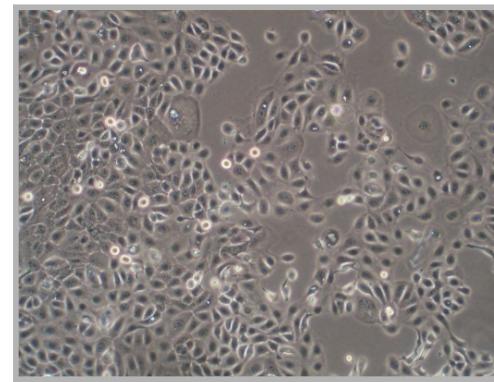
TMA construction

TMA sectioning

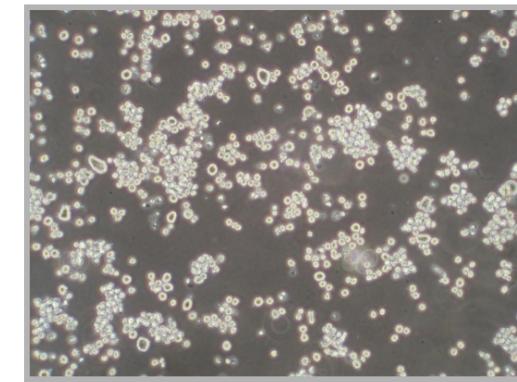


Analysis

# CELL MICROARRAYS

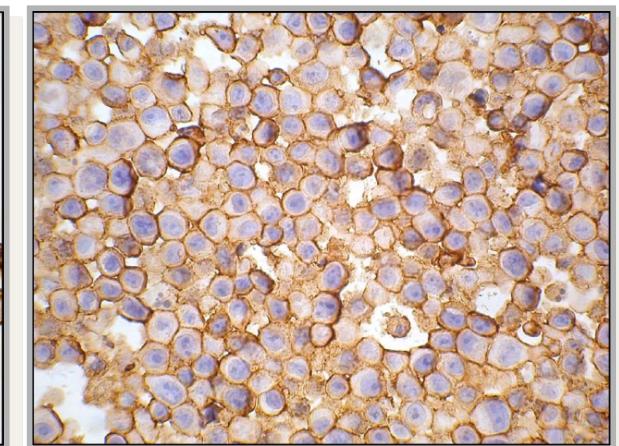


Adherent Cells



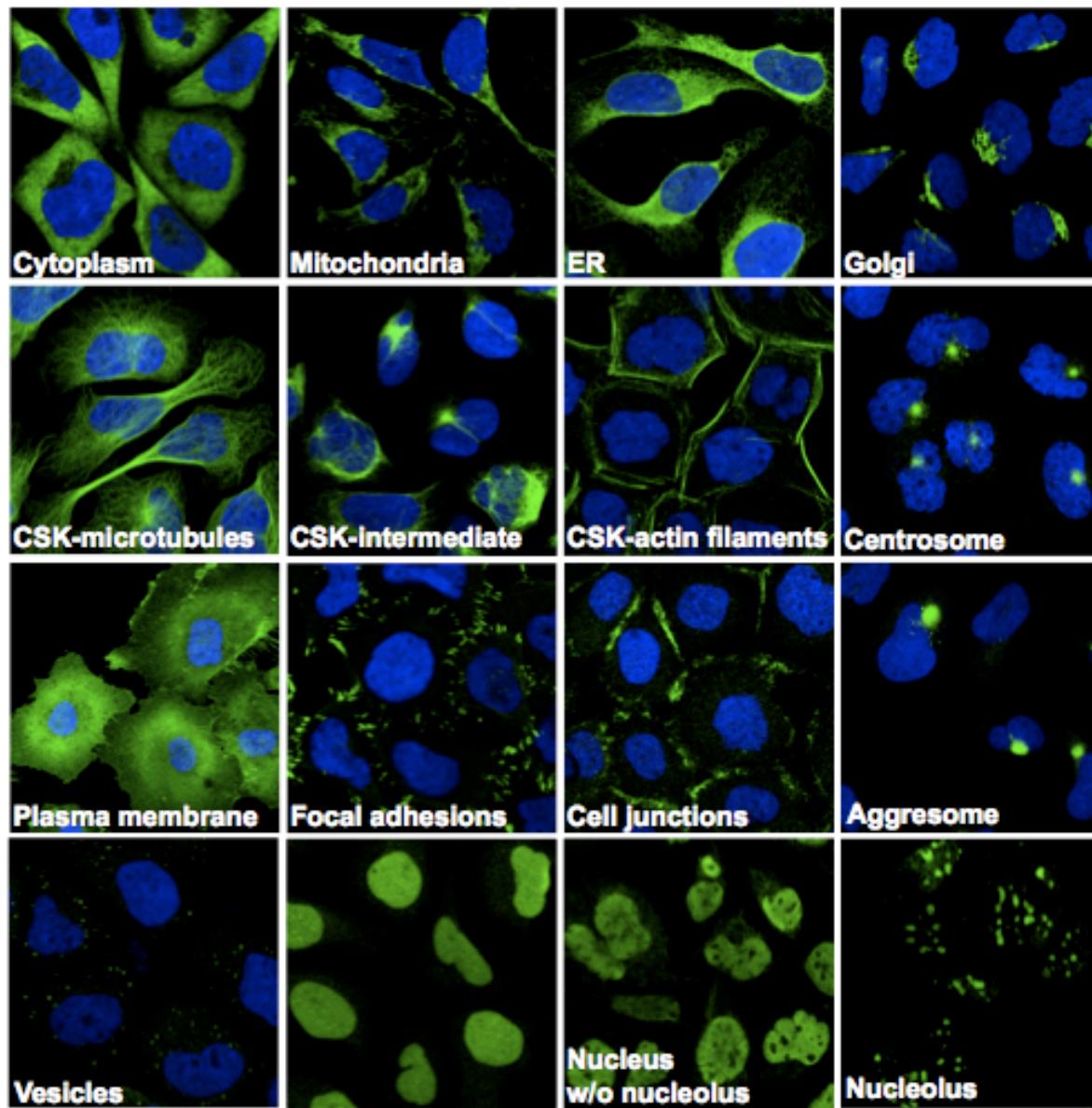
Suspension Cells

- Harvest Cells
- Wash and Fix
- Suspend Agarose
- Histoprocessing

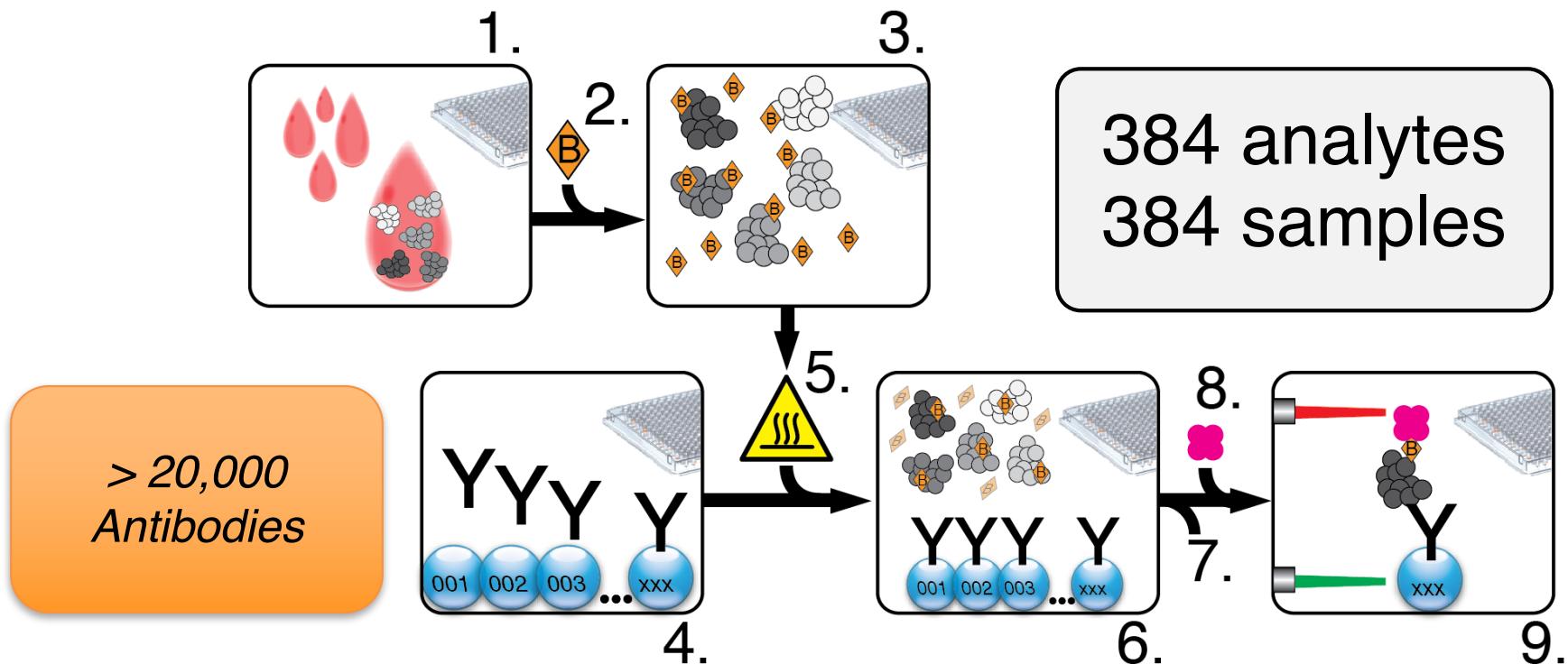


50-100 milj cells/ml agarose

# FLUORESCENCE MICROSCOPY



# ANTIBODY BEAD ARRAYS



# SOME LIMITATIONS OF AFFINITY PROTEOMICS

- Generation and availability of affinity reagents
- Availability of protein of interest
- Selectivity is assay dependent
- Limited quantification possibilities
- Low-throughput (proteins per sample)
- Mostly targeted analysis

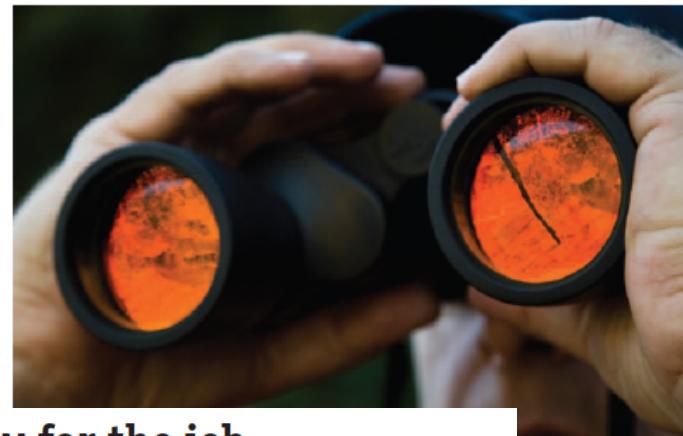
## Calling the next generation of affinity reagents

As complements to antibodies, new reagents to target proteins invite broad types of experiments.

# "ANTIBODIES ARE NOT MAGIC REAGENTS."

274 | NATURE | VOL. 511 | 21 MAY 2015

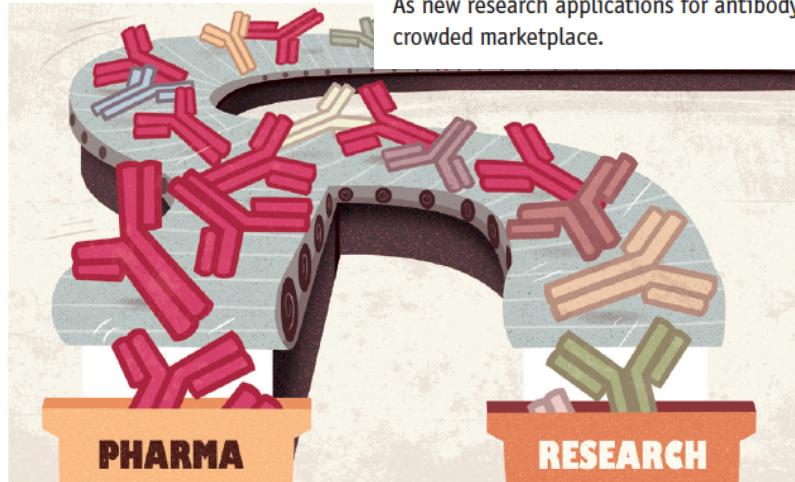
© 2015 Macmillan Publishers Limited. All rights reserved



THINKSTOCK

### Finding the right antibody for the job

Vivien Marx



## Standardize antibodies used in research

To save millions of dollars and dramatically improve reproducibility, protein-binding reagents must be defined by their sequences and produced as recombinant proteins,

5 FEBRUARY 2015 | VOL 518 | NATURE | 27 as Plückthun and 110 co-signatories

## BLAME IT ON THE ANTIBODIES

Antibodies are the workhorses of biological experiments, but they are littering the field with false findings. A few evangelists are pushing for change.

BY MONYA BAKER



# SOME ADVANTAGES OF AFFINITY PROTEOMICS

- Protein localization
- Differentiation of expression sites in tissue
- Applicability across different types of assay
- High-throughput (number of samples)
- High sensitivity assays exist
- Widely used

# MORE INSIGHTS IN COMING LECTURES

- Human Protein Atlas – Peter
- Mass spectrometry – Janne
- Spatial proteomics – Emma
- Tissue proteomics – Jan
- Clinical proteomics – Jacob
- Affinity reagents – Johan
- Affinity coupled mass spectrometry – Claudia
- Plasma profiling with antibodies – workshop

PART 2

# **BIOMARKER DISCOVERY**

# THE HUNT FOR BIOMARKERS

- A biochemical feature that corresponds to a particular physiological state.
- Appear or disappear specifically in disease state.
- Pattern from combination of several biomarkers
- Compare “healthy” and diseased
- “Easy” to discover but very difficult to validate

# TYPES OF BIOMARKERS

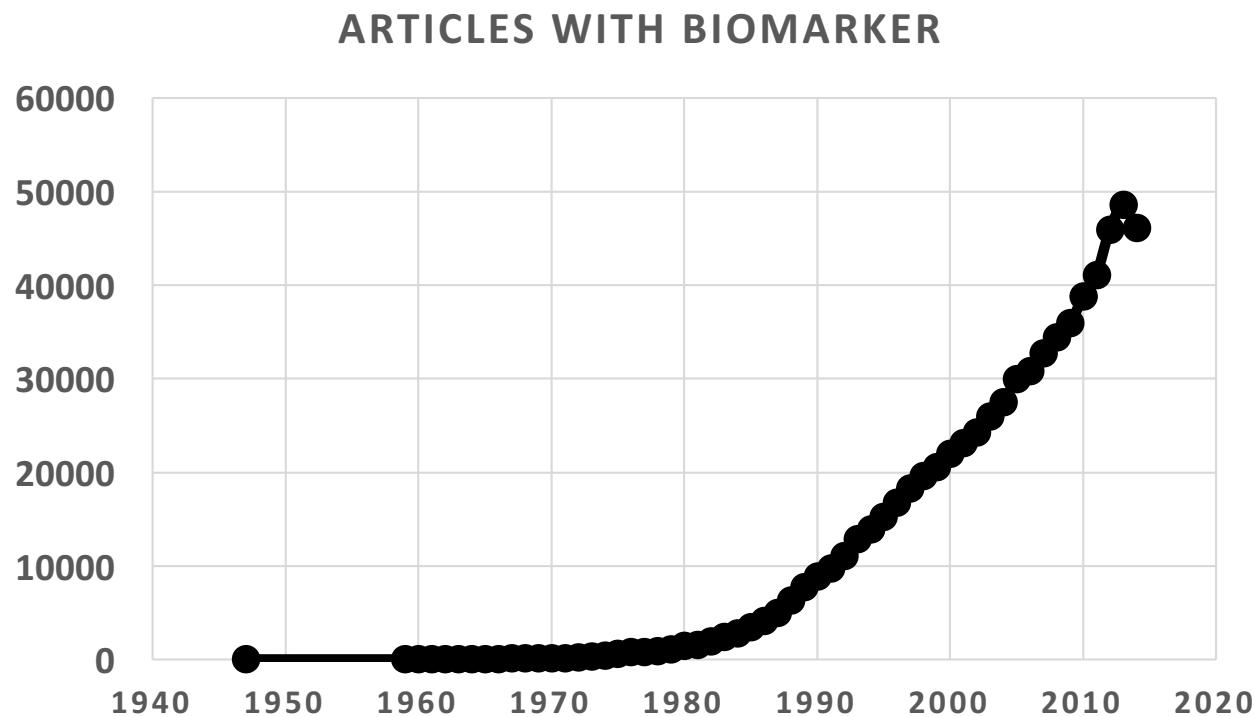
*DIAGNOSTIC:* Status

*PROGNOSTIC:* Risk/Progression

*PREDICTIVE:* Effect/Responds

*...and many more*

**> 710,000 ARTICLES WITH “BIOMARKER”**



# A QUOTE

Despite advances in molecular medicine, genomics, proteomics and translational research, prostate cancer remains the second most common cause of cancer-related mortality for men in the Western world. Clearly, early detection, targeted treatment and post-treatment monitoring are vital tools to combat this disease. Tumor markers can be useful for diagnosis and early detection of cancer, assessment of prognosis, prediction of therapeutic effect and treatment monitoring. Such tumor markers include prostate-specific antigen (prostate), cancer antigen (CA)15.3 (breast), CA125 (ovarian), CA19.9 (gastrointestinal) and serum  $\alpha$ -fetoprotein (testicular cancer). However, all of these biomarkers lack sensitivity and specificity and, therefore, there is a large drive towards proteomic biomarker discovery. Current research efforts are directed towards discovering biosignatures from biological samples using novel proteomic technologies that provide high-throughput, in-depth analysis and quantification of the proteome. Several of these studies have revealed promising biomarkers for use in diagnosis, assessment of prognosis, and targeting treatment of prostate cancer. This review focuses on prostate cancer proteomic biomarker discovery and its future potential.

# CHALLENGE FOR BIOMARKERS

While others are much more aggressive. One of the major hurdles in biomarker discovery is the increasing heterogeneity of the disease as it progresses, to the extent that advanced PCa often represents a group of several diseases, even within the same patient, with differing morphologies, immunophenotypes and genotypes [10].

# DREAM SCENARIO

teomic methods. An ideal tumor marker would have a high sensitivity and specificity, be practical, cost effective and minimally invasive to test [15]. Although many biomarkers exist, none of them fit all of these criteria. Indeed, it may never be possible for one single biomarker to fulfill all of the ideal criteria and, instead it, may be necessary to use panels and series of markers (so-called multiplexing) [29]. It should, however, be remembered that such panels would risk increasing sensitivity at the expense of specificity. Recent advances in high-throughput technologies, in which a large proportion of gene and/or protein expression in a large number of samples can be quickly quantified and compared, help widen the search for such biomarkers [30].

# KEY CHALLENGES IN PROTEOMIC ANALYSIS

## Sensitivity

possibility to correctly detect few copies of a protein

## Robustness

possibility to generate the same data again and again

## Throughput

number of samples and proteins per experiment

**Table 1.** Example sensitivities and specificities for the nine FDA approved cancer biomarkers.

| Marker   | Disease                         | Cut Off                | Sensitivity | Specificity      | Reference                  |
|--|---------------------------------|------------------------|-------------|------------------|----------------------------|
| CEA  | malignant pleural effusion      | NA <sup>1</sup>        | 57.5%       | 78.6%            | (Li et al. 2003)           |
| CEA  | peritoneal cancer dissemination | 0.5 ng/ml              | 75.8%       | 90.8%            | (Yamamoto et al. 2004)     |
| Her-2/neu  | stage IV breast cancer          | 15 ng/mL               | 40%         | 98% <sup>2</sup> | (Cook et al. 2001)         |
| Bladder Tumor Antigen                            | urothelial cell carcinoma       | NA                     | 52.8%       | 70%              | (Mian et al. 2000)         |
| Thyro-globulin                                   | thyroid cancer metastasis       | 2.3 ng/ml <sup>3</sup> | 74.5%       | 95%              | (Lima et al. 2002)         |
| Alpha-fetoprotein                                | hepatocellular carcinoma        | 20 ng/ml               | 50%         | 70%              | (De Masi et al. 2005)      |
| PSA  | prostate cancer                 | 4.0 ng/mL              | 46%         | 91%              | (Gann et al. 1995)         |
| CA 125   | non-small cell lung cancer      | 95 IU/mL               | 84%         | 80%              | (Dabrowska et al. 2004)    |
| CA19.9   | pancreatic cancer               | NA                     | 75%         | 80%              | (Yamaguchi et al. 2004)    |
| CA 15.3  | breast cancer                   | 40 U/ml                | 58.2%       | 96.0%            | (Ciambellotti et al. 1993) |
| leptin, prolactin, osteopontin, and IGF-II       | ovarian cancer                  | NA                     | 95%         | 95%              | (Mor et al. 2005)          |
| CD98, fascin, sIgR <sup>4</sup> , and 14-3-3 eta | lung cancer                     | NA                     | 96%         | 77%              | (Xiao et al. 2005)         |
| Troponin I                                       | myocardial infarction           | 0.1 microg/L           | 93%         | 81%              | (Eggers et al. 2004)       |
| B-type natriuretic peptide                       | Congestive heart failure        | 8 pg/mL                | 98%         | 92%              | (Dao et al. 2001)          |

1. Not available

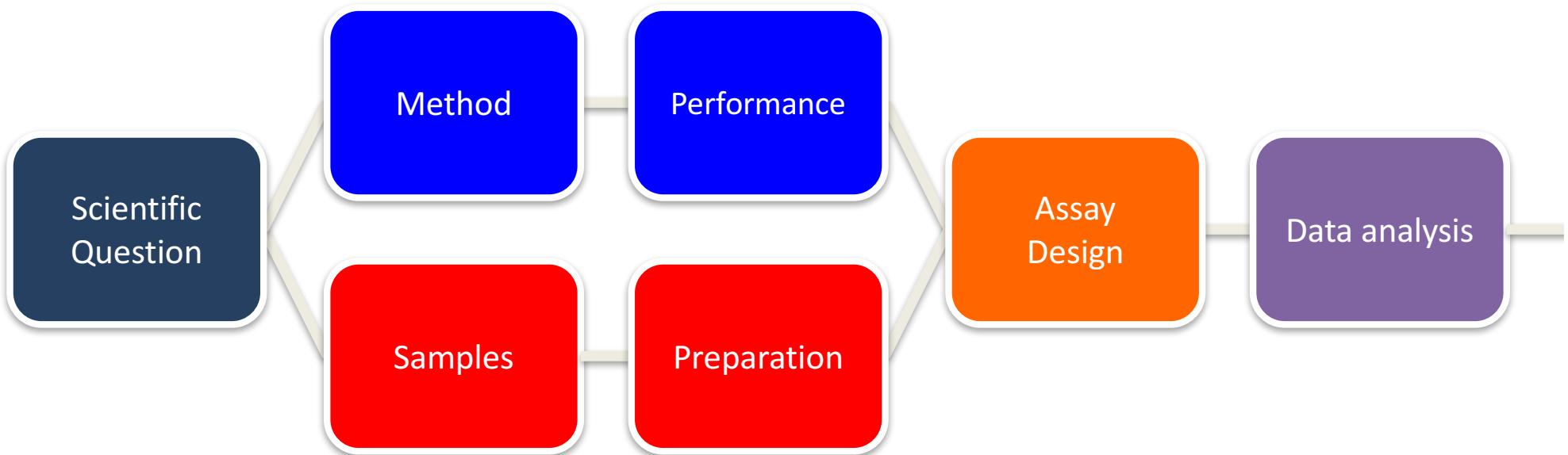
2. vs benign breast diseases

3. vs 3rd week post surgery

4. Secreted chain of the polymeric immunoglobulin receptor

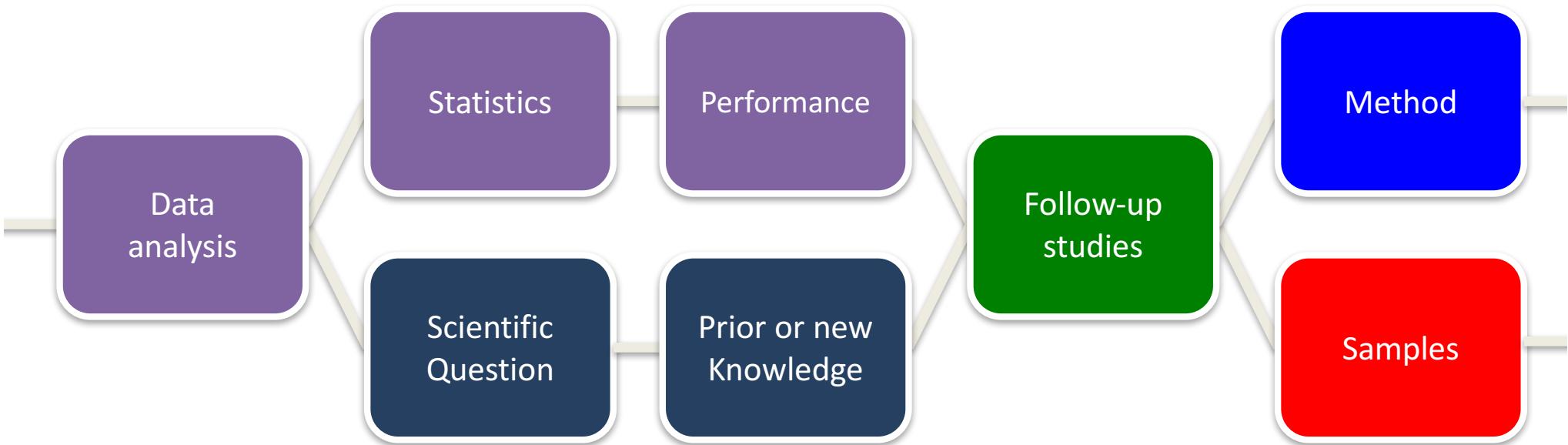
# APPROACH TO BIOMARKER DISCOVERY:

## Experimental design I



# **APPROACH TO BIOMARKER VALIDATION:**

## **Experimental design II**



# **EXAMPLE ON PROSTATE CANCER**

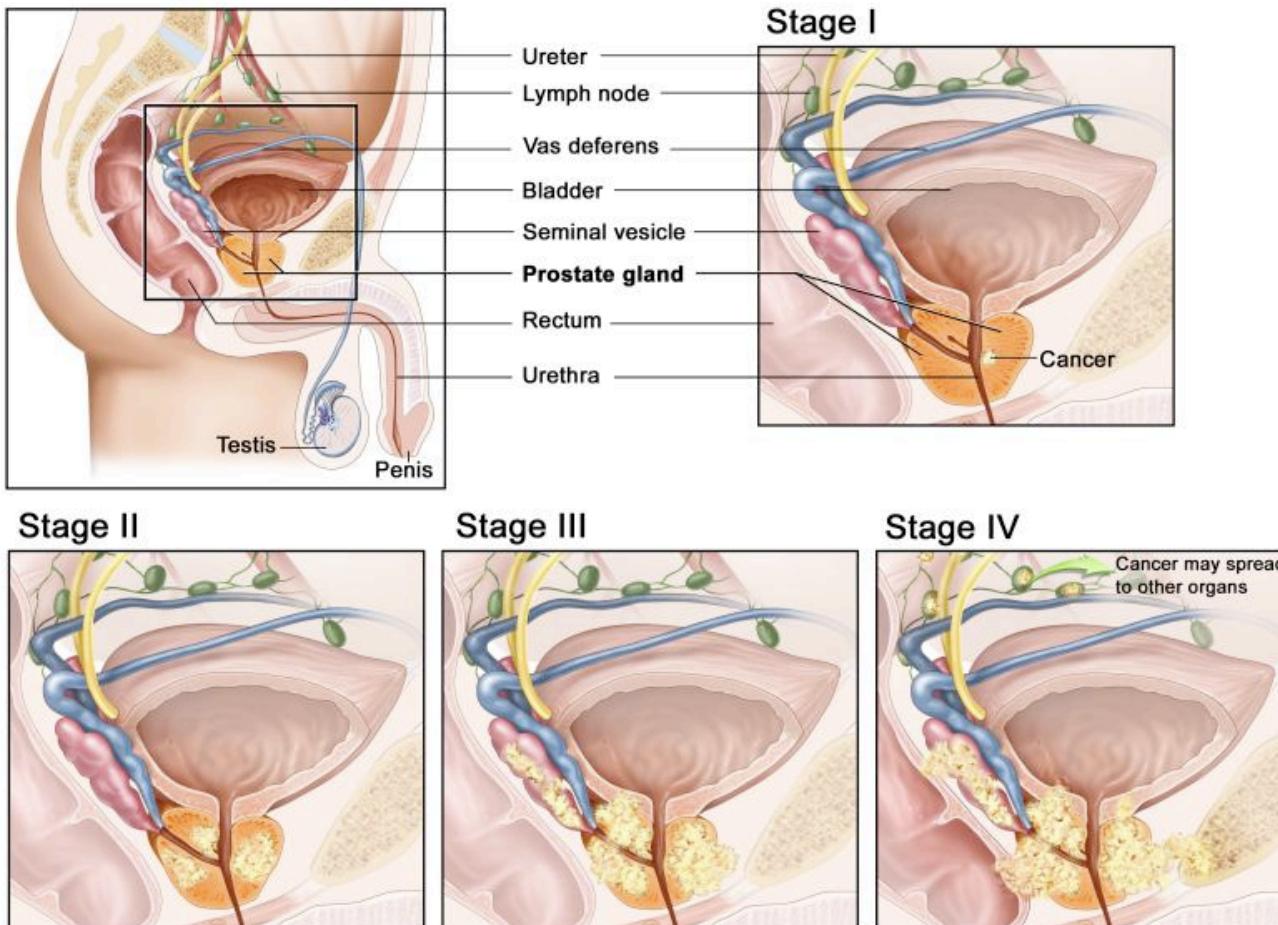
# **WHAT IS CANCER?**

- 
- 
- 
- 
- 
- 
-

# Cancer stages (TNM)

- Primary Tumor (T)
  - TX Primary tumor cannot be evaluated
  - T0 No evidence of primary tumor
  - Tis Carcinoma in situ (early cancer that has not spread to neighboring tissue)
  - T1, T2, T3, T4 Size and/or extent of the primary tumor
- Regional Lymph Nodes (N)
  - NX Regional lymph nodes cannot be evaluated
  - N0 No regional lymph node involvement (no cancer found in lymph nodes)
  - N1, N2, N3 Involvement of regional lymph nodes (number and/or extent of spread)
- Distant Metastasis (M)
  - MX Distant metastasis cannot be evaluated
  - M0 No distant metastasis (cancer has not spread to other parts of the body)
  - M1 Distant metastasis (cancer has spread to distant parts of the body)

# PROSTATE CANCER STAGES



National Cancer Institute

[www.meb.uni-bonn.de/cancer.gov/CDRooooo62965.html](http://www.meb.uni-bonn.de/cancer.gov/CDRooooo62965.html)

# **What can be studied...**

- **Prostate cancer genetics**
  - Genetic Epidemiology
  - Recent development in prostate cancer genetics
- **Discovery of other biomarkers**
  - Proteomics
  - Metabolomics
  - Tumor markers
- **Future use of genetic markers in the clinic**
  - Before diagnosis
  - At time of prostate biopsy
  - At time of treatment decision

# PCA FACTS (US)

- Prostate cancer is the most common cancer behind skin cancer
- About 1 of 7 man will be diagnosed with prostate cancer during his lifetime.
- In the United States an 2015 there are
  - About 220,800 new cases of prostate cancer
  - About 27,540 deaths from prostate cancer
- About 6 cases in 10 are diagnosed in men aged 65 or older, and it is rare before age 40.
- The average age at the time of diagnosis is about 66.
- Prostate cancer is the second leading cause of cancer death in American men, behind only lung cancer. About 1 man in 38 will die of prostate cancer.
- About 60 % of all prostate cancers are diagnosed in men 65 years of age and older.

<http://www.cancer.org/cancer/prostatecancer/>

# KEY QUESTIONS IN PROSTATE CANCER

- Why do men get prostate cancer?
- What risk factors exist?
- When to screening for cancer?
- Can we find markers for aggressive prostate cancer before diagnosis or treatment?

# DIAGNOSTICS

## Diagnostics

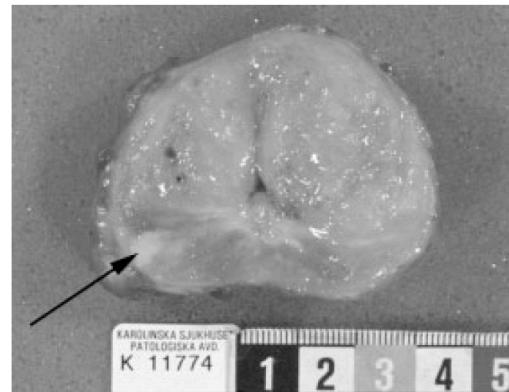
- Ultrasound on prostate tissue
- Seminal fluid
- Urine
- Serum test (PSA)
- Family history

## Treatment

- Surgery
- Chemotherapy
- Hormonal therapy
- Radiation therapy
- Watchful waiting
- Cryosurgery

# PROSTATE TISSUE SAMPLE

- Small organ (size of a walnut)
- Tissue section from biopsy
- Fixation for histological analysis
- Laser capture micro-dissection
- Tissue lysis



**Figure 2.** Horizontal section through radical prostatectomy specimen with small cancer. The tumor appears as a poorly circumscribed patch in the left posterior corner (arrow).

# THE PROSTATE TISSUE PROTEOME

Table 3. Highly and Moderately enriched genes in normal Prostate.

| Gene name    | Description  | Tissue-specific score | Mean Prostate FPKM | Max FPKM in other tissue |
|--------------|--|-----------------------|--------------------|--------------------------|
| KLK3         | kallikrein-related peptidase 3                       | 816.11                | 4700.79            | 5.76                     |
| KLK2         | kallikrein-related peptidase 2                       | 485.25                | 1682.84            | 3.47                     |
| TGM4         | transglutaminase 4 (prostate)                        | 86.13                 | 858.59             | 9.97                     |
| RLN1         | relaxin 1  | 71.53                 | 37.05              | 0.52                     |
| KLK4         | kallikrein-related peptidase 4                       | 58.16                 | 199.03             | 3.42                     |
| ACPP         | acid phosphatase, prostate                           | 52.20                 | 1941.95            | 37.20                    |
| CHRNA2       | cholinergic receptor, nicotinic, alpha 2 (neuronal)  | 47.78                 | 32.15              | 0.67                     |
| SLC45A3      | solute carrier family 45, member 3                   | 17.65                 | 369.97             | 20.96                    |
| SP8          | Sp8 transcription factor                             | 14.22                 | 1.78               | 0.13                     |
| OR51E2       | olfactory receptor, family 51, subfamily E, member 2 | 12.13                 | 48.96              | 4.04                     |
| RFPPL2       | ret finger protein-like 2                            | 12.04                 | 16.53              | 1.37                     |
| RP11-362K2.2 | Uncharacterized protein                              | 10.80                 | 9.79               | 0.91                     |
| STEAP2       | STEAP family member 2, metalloreductase              | 10.55                 | 136.87             | 12.97                    |
| OR51C1P      | olfactory receptor, family 51, subfamily C, member 1 | 9.84                  | 1.74               | 0.18                     |
| NKX3-1       | NK3 homeobox 1                                       | 8.76                  | 229.15             | 26.14                    |
| MSMB         | microseminoprotein, beta-                            | 8.59                  | 3181.86            | 370.44                   |
| NEFH         | neurofilament, heavy polypeptide                     | 8.55                  | 173.22             | 20.26                    |
| POTEM        | POTE ankyrin domain family, member M                 | 5.76                  | 13.63              | 2.37                     |
| TRIM72       | tripartite motif containing 72                       | 5.64                  | 2.76               | 0.49                     |
| RDH11        | retinol dehydrogenase 11 (all-trans/9-cis/11-cis)    | 5.60                  | 517.68             | 92.52                    |
| NCAPD3       | non-SMC condensin II complex, subunit D3             | 5.48                  | 127.43             | 23.25                    |
| LY6G6D       | lymphocyte antigen 6 complex, locus G6D              | 5.42                  | 3.31               | 0.61                     |

doi:10.1371/journal.pone.0133449.t003

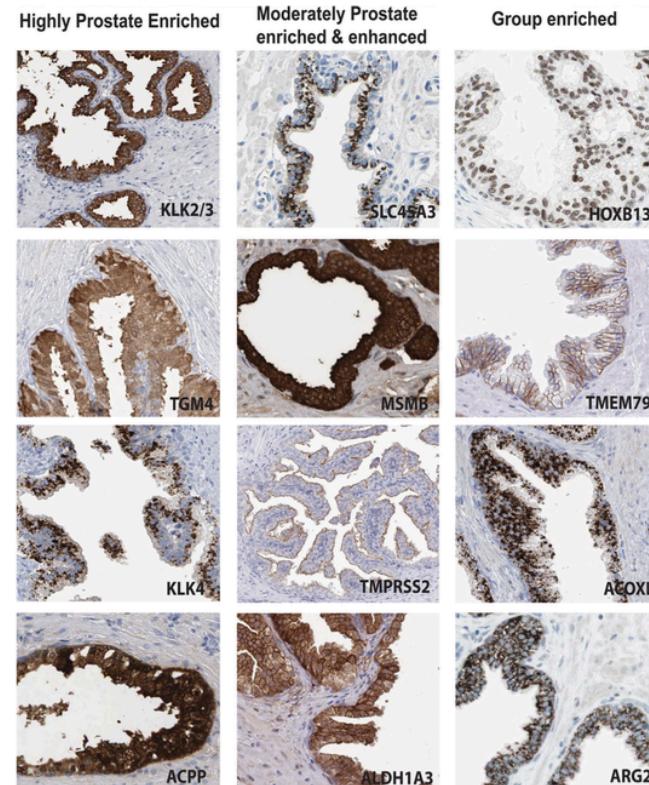
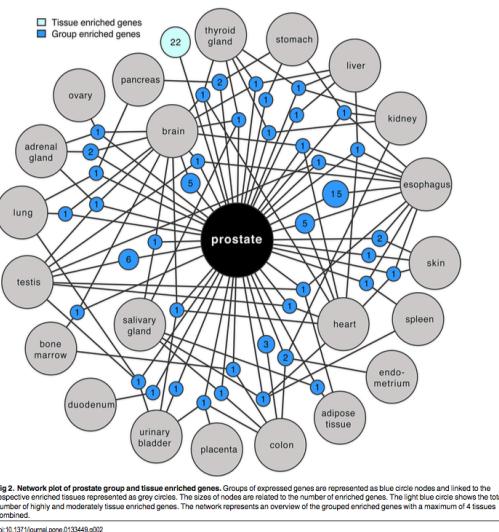


Fig 3. IHC-based protein expression patterns in normal prostate. Examples of expression in benign glandular cells are shown for a subset of proteins corresponding to genes with elevated expression in prostate. Scale, 100 μm.

doi:10.1371/journal.pone.0133449.g003

# **PROXIMAL SAMPLE FLUIDS**

## **URINE**

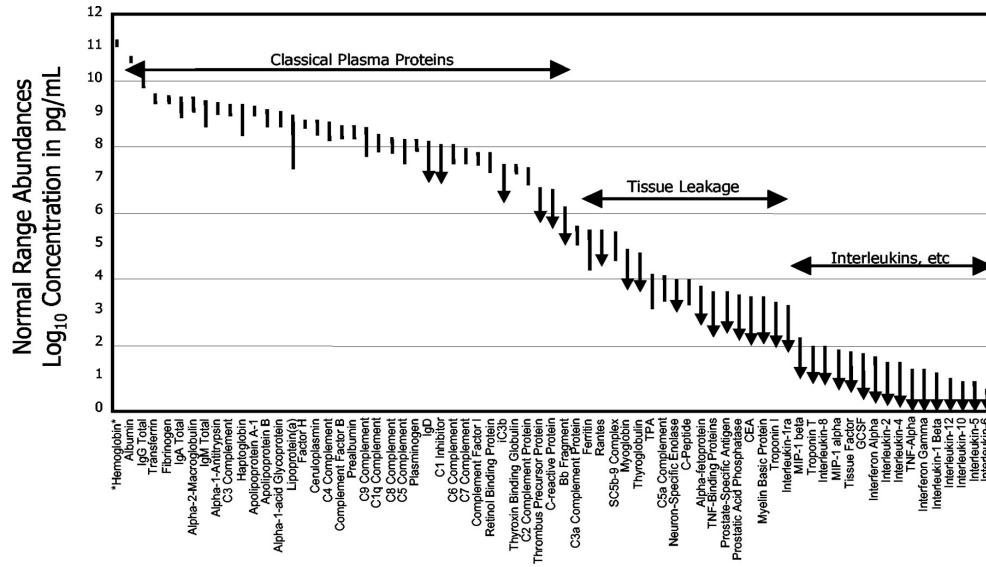
- pH
- Blood
- Glucose
- Low protein content
- Metabolites

## **SEMINAL FLUID**

- Hormone rich
- Growth factors
- Bioreactive peptides

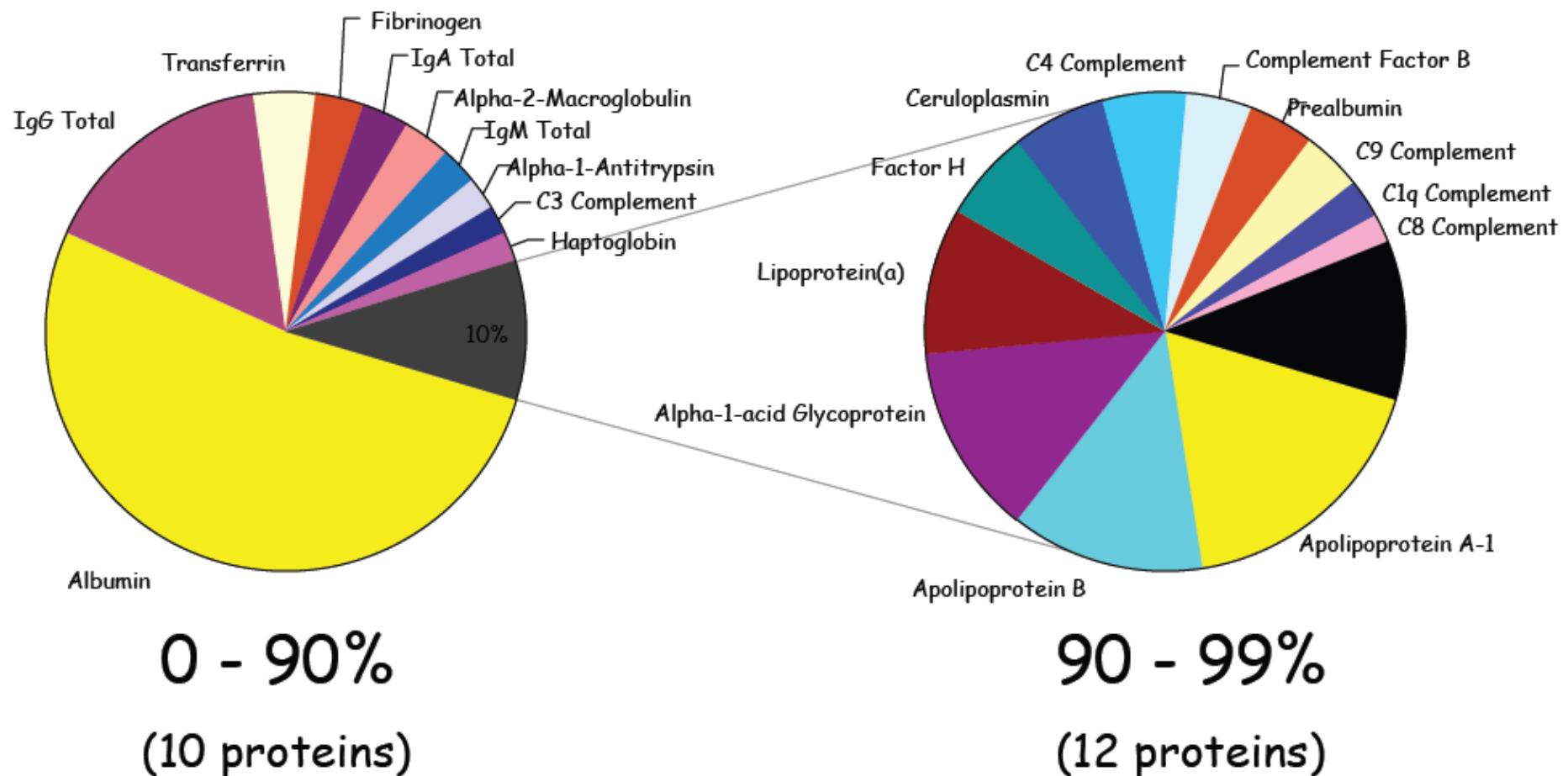
# WHY SERUM AS SOURCE FOR BIOMARKERS?

- Mostly used sample type in clinical diagnostics
  - Minimally-invasive collection procedure
  - Reflection of health status
  - Transportation system



Taken from – Anderson (2002) MCP

# A Small Number of Proteins Make Up the Top 99% of Plasma by Mass



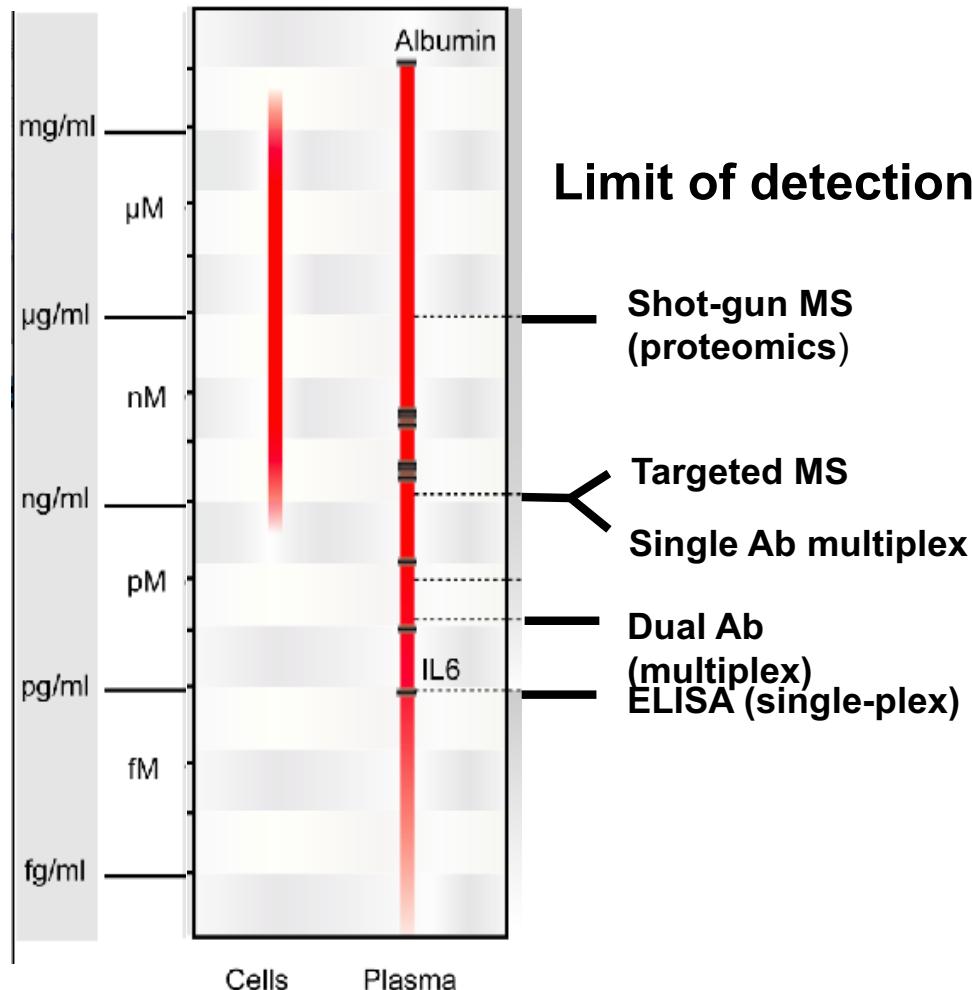
The human plasma proteome: History, character, and diagnostic prospects. Anderson, N.L. and Anderson, N.G., *Molecular and Cellular Proteomics*, 1.11, 845-867 (2002)

# SERUM COMPONENTS

- Secreted proteins acting in plasma
  - *From liver and intestine; kidney filtration cut-off (45 kDa)*
- Immunoglobulin
  - *Unique class, 10<sup>7</sup> different sequences present in circulation*
- “Long Distance” Receptors
  - *Peptide and Protein Hormones*
- “Local” Receptors
  - *Cytokines and short distance mediators*
- Temporary Passengers
  - *Non-hormone Proteins on passage to functional site – e.g. Lysosomal Proteins*
- Tissues Leakage Products
  - *As cause of cell damage or death – Diagnostic markers*
- Aberrant Secretion
  - *From tumors and diseased tissue without functional requirement – Cancer markers*
- Foreign Proteins
  - *Infectious organisms or parasites*



# Protein concentrations in serum



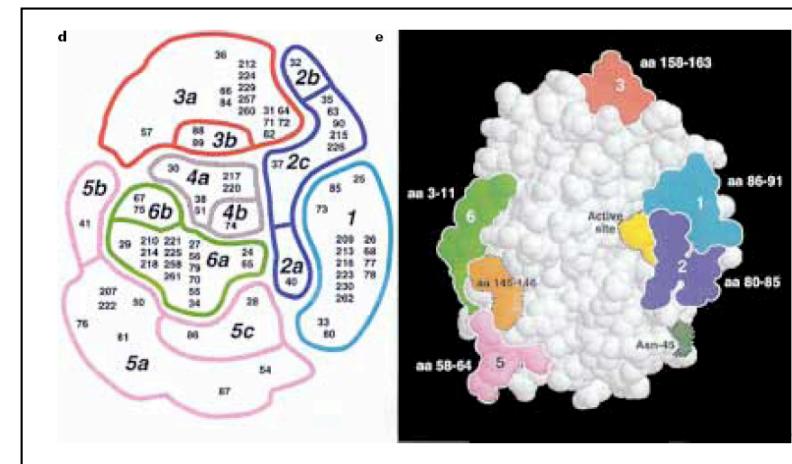
**Limit of detection:**

**Classical proteomics only targets high abundant proteins**

**Dual antibody assays necessary for detection of low abundant proteins in plasma**

# PROSTATE SPECIFIC ANTIGEN (PSA)

- 261 aa, 28,7 kDa
- Glycoprotein
- Secreted
- Serine peptidase S1 family (hydrolysis of semenogelin-1)
- Kallikrein subfamily (KLK3)
- Negative regulation of angiogenesis
- ELISA
- Serum cut-off < 4 ng/ml
- Free and bound PSA



Stenman et al (1999) Tumor Biol

# ISSUES WITH PSA

- + Too unspecific as marker (false positives)
- 

*PSA is prostate specific, but not prostate-cancer specific. The positive predictive value of a PSA > 4.0 ng/mL is only 25% from a pooled meta-analysis of PSA studies.*

---

- + BPH or prostatitis  PSA
- 

*Some investigators question the role of PSA as a screening tool altogether, arguing that it leads to a large number of unnecessary prostate biopsies and probably a large number of potentially unnecessary therapies without significantly impacting on cancer-relative survival.*

---

- Additional biomarkers needed

# NEED FOR MORE BIOMARKER

- Early detection of disease
- Diagnosis at onset
- Tumor initiation and progression
- Clinical outcome/prognosis
- Measure treatment effect
- Assess risk/tumor aggressiveness
- Novel target to therapy

# **CONSIDERATION TO BE MADE**

## **METHOD**

- MS based
  - Targeted or Discovery
  - Mode of detection
  - Quantitative
  - Isoforms or Truncations
  - PTMs
- Antibody based
  - Selected markers
  - Select binding molecules
  - Chose type of assay

## **TARGET**

- Where to be measured
- Function (druggable?)
- Tissue specificity
- Subcellular Location
- Half life
- Interaction partners
- Homology

# **ANOTHER TYPE OF BIOMARKER AUTOANTIBODIES**

# Autoantibodies

- Directed against one or more “own” proteins
- Causing inflammation and damage
- Genetic predisposition combined with environmental trigger as suggested cause
- Hormonal component is discussed

# AUTOANTIBODIES IN PROSTATE CANCER

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

## Autoantibody Signatures in Prostate Cancer

Xiaoju Wang, Ph.D., Jianjun Yu, M.S., Arun Sreekumar, Ph.D.,  
Sooryanarayana Varambally, Ph.D., Ronglai Shen, M.S., Donald Giachiero, Ph.D.,  
Rohit Mehra, M.D., James E. Montie, M.D., Kenneth J. Pienta, M.D.,  
Martin G. Sanda, M.D., Philip W. Kantoff, M.D., Mark A. Rubin, M.D.,  
John T. Wei, M.D., Debasish Ghosh, Ph.D., and Arul M. Chinnaiyan, M.D., Ph.D.

PNAS

CrossMark  
click for updates

## Discovery and horizontal follow-up of an autoantibody signature in human prostate cancer

Paul J. Mintz<sup>a</sup>, Anna Cecilia Rietz<sup>b,c</sup>, Marina Cardó-Vila<sup>b,c</sup>, Michael G. Ozawa<sup>d</sup>, Eleonora Dondossola<sup>d</sup>, Kim-Anh Do<sup>e</sup>, Jeri Kim<sup>a</sup>, Patricia Troncoso<sup>f</sup>, Christopher J. Logothetis<sup>a</sup>, Richard L. Sidman<sup>g,1</sup>, Renata Pasqualini<sup>b,c,1,2</sup>, and Wadih Arap<sup>b,h,1</sup>

Departments of <sup>a</sup>Genitourinary Medical Oncology, <sup>b</sup>Biostatistics, and <sup>c</sup>Pathology and <sup>d</sup>David H. Koch Center, The University of Texas M. D. Anderson Cancer Center, Houston, TX 77030; <sup>e</sup>Harvard Medical School and Department of Neurology, Beth Israel Deaconess Medical Center, Boston, MA 02215; and <sup>f</sup>University of New Mexico Cancer Center and Divisions of <sup>g</sup>Molecular Medicine, and <sup>h</sup>Hematology and Medical Oncology, Department of Internal Medicine, University of New Mexico School of Medicine, Albuquerque, NM 87131

The Prostate 72:427–436 (2012)

## Serum-Autoantibodies for Discovery of Prostate Cancer Specific Biomarkers

Petra Massoner,<sup>1</sup> Angelika Lueking,<sup>2</sup> Heike Goehler,<sup>2</sup> Annabel Höpfner,<sup>2</sup> Axel Kowald,<sup>2</sup> Karl G. Kugler,<sup>3</sup> Peter Amersdorfer,<sup>4</sup> Wolfgang Horninger,<sup>1</sup> Georg Bartsch,<sup>1</sup> Peter Schulz-Knappe,<sup>2</sup> and Helmut Klocker<sup>1\*</sup>

<sup>1</sup>Division of Experimental Urology, Department of Urology, Innsbruck Medical University, Innsbruck, Austria  
<sup>2</sup>Protagen AG, Dortmund, Germany  
<sup>3</sup>Institute for Bioinformatics and Translational Research, University for Health Sciences, Medical Informatics and Technology (UMIT), Hall in Tirol, Austria  
<sup>4</sup>DiagnoNET, Graz, Austria

JOURNAL OF PROTEOMICS 119 (2015) 218–229

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

ScienceDirect

[www.elsevier.com/locate/jprot](http://www.elsevier.com/locate/jprot)

Journal of PROTEOMICS

CrossMark

## Prostate cancer-associated autoantibodies in serum against tumor-associated antigens as potential new biomarkers

Ramesh Ummanni<sup>a,b,\*1</sup>, Divya Duscharla<sup>b,c,1</sup>, Christine Barrett<sup>a</sup>, Simone Venz<sup>d,e</sup>, Thorsten Schlomm<sup>f</sup>, Hans Heinzer<sup>f</sup>, Reinhard Walther<sup>d</sup>, Carsten Bokemeyer<sup>a</sup>, Tim H. Brümmendorf<sup>g</sup>, P.V.L.N. Murthy<sup>h</sup>, Stefan Balabanov<sup>a,i,\*\*</sup>

<sup>a</sup>Department of Oncology, Haematology, Hubertus Wald-Tumour Zentrum (JCCH), University Hospital Eppendorf (UKE), Hamburg, Germany  
<sup>b</sup>Center for Chemical Biology, Indian Institute of Chemical Technology (IITCT), Hyderabad, India  
<sup>c</sup>Centre for Academy of Scientific and Innovative Research, CSIR-Indian Institute of Chemical Technology (CSIR-IICT), Hyderabad, India  
<sup>d</sup>Department of Medical Biochemistry and Molecular Biology, University of Greifswald, Greifswald, Germany  
<sup>e</sup>Interdisciplinary Institute of Genetics and Functional Genomics, University of Greifswald, Greifswald, Germany  
<sup>f</sup>Prostate Cancer Centre, University Hospital Eppendorf (UKE), Hamburg, Germany  
<sup>g</sup>Department of Haematology and Oncology, University Hospital of the RWTH Aachen, Aachen, Germany  
<sup>h</sup>Department of Urology, Nizam's Institute of Medical Sciences (NIMS), Hyderabad, India  
<sup>i</sup>Division of Haematology, University Hospital Zurich, Zurich, Switzerland

NEXT

# THE HUMAN PROTEIN ATLAS